THE UNIVERSITY
of EDINBURGH

Investigating catalysis-independent roles of PRC2

Marie Warburton

Doctor of Philosophy
The University of Edinburgh
School of Biological Sciences
August 2019

Abstract

The catalytic subunit EZH2 of Polycomb Repressive complex 2 (PRC2) is responsible for placing the histone H3 lysine 27 trimethylation (H3K27me3) histone modification at nucleosomes, a mark associated with silenced genes. This histone modification can be found at domains that also carry marks usually associated with a transcriptionally active state, such as H3K4me3 and H3K4me1. Genes that bear this overlap of repressive and active marks around their promoters and/or enhancers are known as bivalent or poised genes, thought to be in a state of readiness for transcriptional silencing or activation upon stimulation by additional factors during differentiation. However, it is not known how much of Polycomb repression relies on H3K27me3 alone and whether the binding of PRC2 itself plays a part that is independent of this histone modification. It has further been suggested that PRC2 may interact with chromatin organisers and be involved in long-range chromatin interactions. Recently, it has also been noted that poised enhancers are already looped to contact their respective promoters, in a PRC2 dependent manner, before gene activation occurs. Again, it is not yet known whether these contacts depend on H3K27me3 or whether they are mediated by the protein complex itself. To address these questions, I used CRISPR/Cas9 gene editing to generate both a knock out and catalytically inactive form of EZH2 in mouse embryonic stem cells, both in the presence and absence of its less active paralogue, EZH1. By performing gene expression, chromatin conformation, and differentiation assays, I aimed to identify roles of PRC2 that may be independent of the histone mark that it catalyses and that may mediate the role of PRC2 in genome organisation, as well as seeking to further elucidate the role of PRC2 throughout development. The findings of this project showed that some catalytic activity of PRC2 was required for the maintenance of Polycomb gene silencing in mESCs. Interestingly, the levels of H3K27me3 that required to maintain repression were much lower than those that are found at these genes in wild-type conditions. I found that not only was the activity of

EZH1 sufficient to maintain the silencing of Polycomb bound genes in mESCs, but also for the majority of RING1b recruitment to these sites, and the maintenance of the PRC2-dependent enhancer-promoter contact at the *Lhx5* locus. Furthermore, I found that throughout differentiation EZH1 was able to maintain the repression of genes that were silenced in undifferentiated mESCs, but EZH2 and robust H3K27me3 methylation was required to allow for the correct upregulation of certain developmental genes required for transition to the neural lineage. This failure to upregulate such a large number of genes was somewhat unexpected as PRC2 is known as a repressive complex and this may suggest it also possesses some gene activating functions. Based on these findings I concluded that while the low catalytic activity of EZH1 is sufficient to maintain a Polycomb repressive state at target genes in a steady state environment such as undifferentiated mESCs, the activity of EZH2 is required in the dynamic conditions of cell differentiation to establish robust levels of H3K27me3 deposition and Polycomb binding at novel target sites.

Lay Summary

One of the biggest questions that has interested biologists since the earliest days of research in this field is how a complex adult individual composed of a multitude of highly specialised organs and tissues can arise from a single cell. As this single cell divides and develops into an embryo, every subsequent cell that is generated contains the exact same genetic information, and yet these cells can go on to fulfil an extremely diverse set of functions. This is achieved by precisely controlling the subset of genes that are active, or expressed, in each cell, giving it its distinct characteristics within the organism. Each cell in the embryo becomes more and more specialised throughout development as the set of genes it expresses becomes more and more selective and fine-tuned. The process by which this happens is still not fully understood, but it relies on many different layers of regulation. One such layer involved the structure that is formed by the DNA and proteins known as histones. DNA is the molecule within the cell that carries all the genetic information. This is an extremely long molecule and is contained within the nucleus of the cell by being packaged in an efficient and organised way. This packaging involves the aforementioned histones, which form a globular structure, around which the DNA can be wound, forming what is known as a nucleosome. These nucleosomes form across the length of the DNA resulting in what is termed the beads on a string structure of the chromatin fibre. This string of nucleosomes can then coil together into a condensed and regular structure enabling the DNA to be efficiently packed into a small space. The state of this chromatin around the position of a given gene has a strong effect on whether this gene is active or not. If the chromatin is tightly packed with the nucleosomes being very close together across the gene, it is more difficult for this gene to be expressed. If the chromatin is more loosely packed and the DNA becomes more accessible with more widely spaced nucleosomes, then the gene will be much more readily expressed. An additional factor governing this process is the state of the histones themselves. These proteins can be modified by the addition of small molecules that can act

as signals or markers of the state of gene expression in that region, among other things. These small molecules are placed onto the histones by a large family of proteins that are collectively known as chromatin modifiers.

My PhD focuses on one of these chromatin modifiers, which is responsible for placing a histone mark that is usually associated with an inactive state of gene expression. This chromatin modifier, known as PRC2, is a complex of multiple separate proteins that have to fit together into a complex in order to be able to place its mark on histones. PRC2 is important for acquiring the correct patterns of gene expression during embryo development and in stem cells, a type of cell that has not yet become specialised, it helps maintain this unspecialised state by keeping genes involved in the development of an adult organism in an inactive state. There are two main types of PRC2 that exist in the cell, one that places high levels of its mark on histones, and one that is less active and only capable of placing only very low amounts of the mark.

By studying PRC2 and its interaction with chromatin, as well as the gene expression patterns in stem cells, I aimed to increase our knowledge of how this complex functions and how much of its role in cells relies of the mark that it places on histones. I also sought to further our understanding of what PRC2 is required for during development, by studying its influence on gene expression throughout the stages of differentiation that occur as stem cells progressively transition into specialised nerve cells. I found that although PRC2 does need to be able to place its histone mark to be able to have an influence on gene expression, a surprisingly small amount of this mark appears to be sufficient for this effect in stem cells. I also showed that although the less active type of PRC2 is able to maintain most of the complex's function in stem cells, as these cells transition into nerve cells the more active PRC2 is needed for the activation of genes that facilitate this transition.

List of abbreviations

µg	Micrograms
µl	Microlitres
3D	Three dimensional
4C	Circular chromosome conformation capture
5-mC	5' methylcytosine
AEBP2	Adipocyte enhancer-binding protein 2
bp	Base pairs
BSA	Bovine serum albumin
CBP	CREB-binding protein
CBX	Chromobox
Cdkn2A	Cycline-dependent kinase inhibitor 2A
cDNA	Complementary DNA
CFP1	CxxC finger protein 1
CGI	CpG Island
ChIP	Chromatin Immunoprecipitation
CNS	Central nervous system
Col1a1/2	Collagen type 1 alpha 1/2
CpG	5'-C-phosphate-G-3'
CRABP1/2	Cellular retinoic acid binding protein 1/2
DAPI	4',6-diamidino-2-phenylindole
dKO	Double Knock-out
DMEM	Dulbecco's modified eagle's medium
DNA	Deoxyribonucleic acid
Dnmt	DNA methyltransferase
DTT	Dithiothreitol
dsDNA	Double stranded deoxyribonucleic acid
ssDNA	Single stranded deoxyribonucleic acid
EB	Embryoid body

EDTA	Ethylenediaminetetraacetic acid
EED	Embryonic Ectoderm Development
ESC	Embryonic stem cell
EZH1	Enhancer of Zeste 1
EZH2	Enhancer of Zeste 2
FABP7	Fatty acid binding protein 7
FACS	Fluorescence-activated cell sorting
FBS	Foetal bovine serum
FGF	Fibroblast growth factor
FISH	Fluorescence in situ hybridisation
FOXA1/2	Forkhead box A1/2
GABA	Gamma-aminobutyric acid
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
gDNA	Genomic DNA
GFP	Green fluorescent protein
GLAST	Glutamate aspartate transporter
GO	Gene ontology
H2A	Histone 2A
H2AK119Ub	Histone 2A Lysine 119 Ubiquitination
H2B	Histone 2B
H3	Histone 3
H3K4me3	Histone 3 Lysine 4 trimethylation
H3K9me3	Histone 3 Lysine 9 trimethylation
H3K27ac	Histone 3 Lysine 27 acteylation
H3K27me3	Histone 3 Lysine 27 trimethylation
H4	Histone 4
HCl	Hydrochloric acid
HOTAIR	HOX transcript antisense RNA
Hox	Homeotic genes
IP	Immuno-precipitation
JARID2	Jumonji and AT-rich interaction domain containing 2
Kb	Kilobases
KDM2B	Lysine-specific demethylase 2B
KLF4	Krppel-like factor 4
LIF	Leukaemia inhibitory factor
Lhx5	LIM Homeobox 5
M	Molar

MBD	Methyl-CpG binding domain
mESCs	Mouse embryonic stem cells
mg	Milligrams
mRNA	Messenger RNA
NaOH	Sodium hydroxide
ncRNA	Non-coding RNA
NEB	New England Biolabs
NPCs	Neuro-progenitor cells
OCT4	Octamer-binding transcription factor 4
OTX2	Orthodenticle homeobox 2
PAX6	Paired box 6
PCA	Principle component analysis
PDK1	3-phosphoinositide-dependent protein kinase-1
PDL	Poly-D-lysine
PBS	Phosphate buffered saline
PcG	Polycomb group
PCGF	Polycomb group RING finger protein
PCR	Polymerase chain reaction
pFA	Paraformaldehyde
PHC	Polyhomeotic
PRC	Polycomb repressive complex
PRDM14	PR domain zinc finger protein 14
PREs	Polycomb response elements
PTMs	Post-translational modifications
qRT-PCR	Quantitative reverse transcriptase polymerase chain reaction
RA	Retinoic acid
RAR	Retinoic acid receptor
RXR	Retinoid X receptor
RING1A/B	Really interesting new gene 1 A/B
RNA	Ribonucleic acid
RT	Reverse transcriptase
RYBP	RING1 and YY1 binding protein
SAM	S-adenosyl methionine
SDS	Sodium dodecyl sulfate
Sgk3	Serine-threonine protein kinase 3
ssODN	Single strand oligodeoxynucleotide
SOX2	SRY-box 2

SUZ12	Suppressor of Zeste 12 protein homolog
TE	Tris EDTA
TBE	Tris borate EDTA
TBX3	T-box 3
TF	Transcription factor
Trim28	Tripartite motif-containing 28
TrkB	Tyrosine receptor kinase B
TSS	Trasncription start site
TrxG	Trithorax group
TUJ1	Class III beta tubulin
VGAT	Vesicular GABA transporter
VGLUT	Vesicular glutamate transporter
WT	Wild type
YY1	Yin yang 1

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

(*Marie Warburton, August 2019*)

Acknowledgements

There are many people without whom this thesis simply would not have happened and who deserve a mention here.

Fist, a huge thank you to everyone in the lab for supporting me, for all of your help particularly with CRISPR and cell culture, for making long days at the bench a lot less lonely, and for always putting the sweets you brought back from your holidays right next to my desk. I literally could not have done this without all of you.

Thank you to the Bird, Buonomo and Wilson labs for your honest and insightful critique and suggestions during lab meetings, and to Nadina specifically for all the helpful discussions and encouragement.
To Philipp of course, thank you for taking me on as a student, helping me develop as a researcher and a person, and never making me afraid to ask a stupid question. Thanks also to Shaun for all your help with the RNA-seq and ChIP-seq data. I dumped so much on your plate and would have been completely lost without your expertise, thanks for helping me make sense of it all.

To all my friends within the centre, thank you for making time between experiments a lot more fun and brightening my days.
Thank you to my Mum and Dad for encouraging and supporting me when I decided on a whim to move to Scotland and for always encouraging my curiosity and creativity. I'm sorry I'm so bad at remembering to call.
To Chris, thank you for always being there, for Saturday morning coffee, for introducing me to podcasts that fuelled many of my busiest days, and for fixing LaTeX for me when I broke it.

And finally to family and friends I haven't seen or talked to in ages, sorry I'm so bad at keeping in touch, this is what I was so busy doing.

Contents

Chapter 1

Introduction

All living organisms rely on regulation of the expression of their genes in order to maintain everyday cellular function, respond to external and internal stimuli and, in the case of more complex organisms, undergo the transition from a single-cell zygote to a fully developed individual. While all cells in a multi-cellular organism have the same DNA sequence and set of genes at their disposal, a vast diversity of different patterns of expression of these genes arises during development from the single-cell zygote onwards, allowing for the specification of different cell types required to perform a plethora of different functions. This divergence of cell types throughout development eventually came to be understood to be based on changes of gene expression that were clonally heritable and did not rely on alterations in the DNA sequence. The regulation of gene expression throughout these processes is controlled by a number of different factors from signalling cascades and transcription factors, the chemical and physical characteristics of the chromatin fibre down to the DNA sequence itself. These factors work in combination to form highly complex regulatory networks that fine tune gene expression patterns for the specific needs of each cell type.

These systems of gene regulation are often associated with epigenetics. The field of epigenetics arose as a branch of research dedicated to elucidating the relationship between genotype and phenotype. The first definition was provided by Waddington who coined the term in the 1940's, and established epigenetics as "the branch of biology that studies the causal interactions between genes and their products which bring the phenotype into being" (Waddington, 1968; Waddington, 2012; Dupont et al., 2009). Much work has been done in this field since its inception, and our understanding of these systems has been further refined giving

rise to today's revised definition of epigenetics as: "the study of mitotically and/or meiotically heritable changes in gene function that cannot be explained by changes in DNA sequence" (Riggs and Porter 1996, (Felsenfeld, 2014)) with the nuance that the newly established epigenetic state should arise from a transient mechanism that is distinct from that required for the maintenance of this state (Berger et al., 2009).

To date many molecular mechanisms have been identified as being involved in this epigenetic regulation of gene expression, most of which involve chemical modifications of DNA or changes to the highly complex DNA-protein structure known as chromatin. In recent years, the term epigenetic has partially fallen out of use in discussions concerning systems of gene regulation, particularly those involving the chemical modification of the chromatin proteins known as histones, as many of them have been found not to be stably inherited.

1.1 Chromatin and histone modifications

1.1.1 Chromatin fibre

Chromatin is the three-dimensional molecular structure made up of DNA and protein that governs packaging and spatial organisation of the genome in the nucleus. The proteins associated with DNA to create chromatin are histones. Two copies of each of the four core histones, H2A, H2B, H3 and H4, assemble into a globular octamer around which 147 base pairs (bp) of DNA wind to create a nucleosome (Burlingame et al., 1985; Luger et al., 1997). An additional histone, H1, known as the linker histone plays a crucial role in the integrity of nucleosomes and the higher order chromatin structure itself. This histone is larger than the core histones and binds to the nucleosome at the point of entry and exit of the DNA that is wound around the core particle, in contact with both the core histones and the DNA. The presence of H1 on nucleosomes facilitates the compaction of chromatin into an organised structure that forms a condensed fibre with a diameter of 30 nm (Robinson and Rhodes, 2006). This 30 nm fibre can additionally form 300 nm loops which further package chromatin into a 250 nm fibre, which itself can be coiled together to eventually form the highly condensed structures of the chromatids which come together in pairs to form a mitotic chromosome (Németh and Längst, 2004). Both the DNA and histones can serve

as a platform for the binding of a variety of factors involved in various functions such as remodelling of chromatin structure, regulation of gene expression, and DNA replication and repair. The binding of these factors can be modulated by the deposition of small molecular groups on either DNA or histones, by enzymes known as chromatin modifiers. These modifications can also serve to change the physical properties of the chromatin fibre and in this way influence gene expression.

1.1.2 DNA and Histone Post-translational modifications

Although the chemical modification of all bases of DNA is theoretically possible, the only modified bases that have been discovered to date are methylated cytosine and adenine (Kumar et al., 2018). Of the two, methylated adenine, or 6-mA, is the least well characterised. It was first discovered in bacteria and has recently been suggested to also exist in eukaryotes, although study of the function of this modification in these organisms remains in the early stages (Kumar et al., 2018). Methylated cytosine, or 5-mC, on the other hand has been extensively studied and has been known to exist for many years. 5-mC is most often found in a CG DNA sequence context in animals, in fact around 80% of all CGs in the mammalian genomes are known to be methylated (Jabbari and Bernardi, 2004). The presence of this modification influences the accessibility of a given region to the binding of factors involved in gene regulation and, if found around promoter or enhancer regions, is usually associated with a silenced state of transcription (Dor and Cedar, 2018; Kumar et al., 2018). DNA methylation is one of the few truly epigenetic modes of regualtion as this mark is stably inherited throughout cell divisions.

The addition of small molecular groups at specific residues is known as the post-translational modification of histones. These post-translational modifications (PTMs) are one of the most important factors governing chromatin state and one of their main functions appears to be the regulation of gene expression. Their influence on transcription is understood to be mainly conveyed by modulating chromatin structure and recruiting or inhibiting the binding of effector proteins (Bannister and Kouzarides, 2011; Taverna et al., 2007). These histone PTMs have been identified at around 60 specific residues both on the globular domain that is most tightly associated with the other histones of the nucleosome, as well as on the more flexible and basic N-terminal tail which protrudes from the core of

the nucleosome and is more readily accessible to interact with additional factors. Several different molecular groups have been identified as being added to histones, acetyl and mono-, di- and tri-methyl groups are perhaps the most extensively studied, but additional residues have also been shown to be phosphorylated, ubiquitylated or sumoylated (Kouzarides, 2007).

Many of these chemical modifications of the chromatin have been shown to be associated with specific genomic regions, often correlating with the transcriptional state. For instance, euchromatic loci are associated with active marks such as mono-, di-, and tri-methylation of histone H3 lysine 4 (H3K4me1/2/3), H3K36me3, acetylation of H3K9 (H3K9ac) and H3K27 (H3K27ac). An example of this is the enrichment of H3K27ac and H3K4me1 that is found across enhancers of active genes (Zentner et al., 2011). Heterochromatic loci are characterised by repressive marks such as H3K9me2/3, H3K27me3, H4K20me2/3 and monoubiquitination of H2A lysine 119 (H2AK119ub) (Bannister and Kouzarides, 2011; Kouzarides, 2007). This type of chromatin can be further sub-categorised into facultative and constitutive heterochromatin, each bearing a different set of these histone modifications with H3K27me3 and H2AK119ub marking the former and H3K9me2/3 the latter. Although we understand the genome-wide distribution of many histone modifications and some of their functions, we are far from uncovering all roles for these modifications individually and especially in the different combinations that can be found at nucleosomes. As specific genomic features were found to often bear a specific set of different histone PTMs, it was proposed that they could combine to generate a "histone code", such that the chromatin state at a given locus is determined at least in part by the combination of PTMs on the nucleosomes at this locus (Jenuwein and Allis, 2001). It should be noted, however, that the term "histone code" has fallen out of use in recent years as it conjures an overly simplistic image of how combinations of histone modifications might function as part of the complex local chromatin landscape (Rando, 2012).

1.1.3 Readers and Writers

The majority of histone PTMs are known to function through their interaction with additional factors. These PTMs influence chromatin in an indirect fashion, by facilitating or obstructing the binding of non-histone proteins. While the complexes that place histone modifications can be termed writers, these

proteins are the readers of histone PTMs. These readers all contain domains that allow them to recognise histone modifications. Acetylation is recognised by bromodomains, methylation by chromodomains and PHD domains, and phosphorylation by 14-3-3 proteins (Kouzarides, 2007; Ruthenburg et al., 2007). The recognition of these marks can lead to the recruitment of these effector proteins which are then able to carry out further modifications of chromatin, activate or repress transcription, or even act as platforms for the recruitment of additional proteins. An example of this is the ING family of proteins, which once bound to H3K4me3 are then able to recruit histone acetyltransferases or histone deacetylases to chromatin (Champagne and Kutateladze, 2009). In contrast to this, some proteins find their recruitment to be blocked by the presence of certain histone modifications, which in turn will act to inhibit their activity on chromatin. For instance the NURD complex associated with gene repression is unable to be recruited to chromatin in the presence of H3K4me3 (Zegerman et al., 2002). DNA methylation mediates its effect on gene regulation in a similar way. Proteins containing methyl-CpG-binding domains (MBD) are able to bind methylated DNA, whereas only unmethylated CpG islands can be bound by CxxC domains (Long et al., 2013; Du et al., 2015). In line with the co-occurrence of multiple different modifications in the same region, many of the reader proteins are able to interact with more than one modification, which may allow for the modulation of the activity of these proteins, either inhibiting or stimulating it with the presence or absence of a specific set of modifications (Ruthenburg et al., 2007; Kouzarides, 2007).

There is a subset of histone PTMs that are known to influence regulation by physically altering chromatin. This type of histone modification functions by creating changes in charge which can affect histone interactions with DNA and disrupt contacts between nucleosomes. An example of this is lysine acetylation. This modification neutralises the positive charge of the lysine, which in turn affects the interaction of the nucleosome with the negatively charged DNA and leads to destabilisation of this structure and a less compact state of chromatin (Shogren-Knaak, 2006). Histone phosphorylation, on the other hand, has been shown in one instance to lead to an increase in condensation of chromatin (Wei et al., 1998).

The "writers" of histone modifications fall into multiple families of protein complexes, which are often themselves able to detect the presence or absence of a certain number of histone marks, which in turn can either stimulate or

attenuate their catalytic activity towards their target residue. Acetylation marks are placed by proteins of the MYST, GNAT and CBP families, ubiquitiylation is placed by RING finger proteins and methylation is placed by the large family of proteins containing SET domains (Kouzarides, 2007). Two major groups of the SET domain containing methyltransferases are the Trithorax and Polycomb families of proteins.

1.2 Polycomb and Trithorax group proteins

Trithorax (TrxG) and Polycomb (PcG) group proteins are chromatin modifiers that play antagonistic roles in gene regulation and have been found to play important roles throughout development. Polycomb group proteins were first discovered in *Drosophila* as regulators of homeotic (Hox) genes (Lewis, 1978). The regulation of transcriptional silencing of these genes was found to play a major role in the body patterning of these organisms by spatially and temporally restricting the expression of key target genes throughout development (Zink and Paro, 1989). PcG proteins assemble to form several complexes, of which there are 2 main families found in mammals, PRC1 and PRC2. These complexes are both chromatin binders and are responsible for placing post-translational modifications on histones H3 and H2A, with PRC1 placing ubiquitylation on lysine 119 of histone H2A, and PRC2 placing methylation on lysine 27 of histone H3. Gene repression is thought to be mediated by these 2 complexes, often in a cooperative manner. The Trithorax group proteins were discovered in parallel with the Polycomb as antagonists to the regulatory functions carried out by the PcG proteins. The main histone PTM associated with TrxG proteins is H3K4me3 and in mammals the primary methyltransferase responsible for its deposition is the SET1A/B complex (Shilatifard, 2012). H3K4me3 is associated with active transcription and is most often found at promoters and transcriptional start sites (TSS) of active genes. This mark is mostly known to function through the recruitment of proteins that promote transcription, such as initiation factor TFIID (Vermeulen et al., 2007; Lauberth et al., 2013), as well as recruiting additional chromatin modifying complexes such as the SAGA histone acetyltransferase complex (Ringel et al., 2015). As well as the SET1A/B complexes, H3K4me3 can additionally be catalysed by the MLL1/2 group of Trithorax proteins. Both these complexes contain CxxC domains that allow them to recognise unmethylated CpG islands, many of which are found at gene

promoters (Birke, 2002; Bach et al., 2009). The MLL3/4 complexes, which also belong to the Trithorax group but do not possess CxxC domains, are responsible for placing H3K4me1 which is mostly enriched at active enhancers (Herz et al., 2012).

As mentioned above, these two families of complexes have mostly antagonistic functions, however are both preferentially targeted to unmethylated CpG islands and as such occasionally target the same domains. This co-localisation will be discussed in a section below (see 1.4).

1.2.1 PRC2

Polycomb repressive complex 2 (PRC2) is conserved in all eukaryotes and composed of 4 core subunits that together form a complex of around 250 kDa responsible for methylating lysine 27 of histone H3 (H3K27), a histone modification associated with gene repression (Kuzmichev, 2002; Czermin et al., 2002; Müller et al., 2002). Of these core subunits, the catalytic activity is conveyed by the SET domain containing EZH1/2 protein which requires both the EED and SUZ12 subunits to be enzymatically active (Cao et al., 2002; Yuan et al., 2012), with further stimulation provided by the RbAp46/48 subunit (Pasini et al., 2004; Ketel et al., 2005).

Non-catalytic core subunits

The Embryonic Ectoderm Development (EED) subunit contains a WD40 domain, which forms an aromatic cage enabling it to bind the H3K27me3 mark placed by PRC2 (Margueron et al., 2009; Poepsel et al., 2018). The binding of this mark by EED leads to the stabilisation of the SRM domain of EZH2 triggering a conformational change of the catalytic subunit to a more active state and thus contributing to the propagation of the histone modification (Margueron et al., 2009; Jiao and Liu, 2015; Justin et al., 2016; Poepsel et al., 2018). This positive feedback loop not only allows for the propagation of H3K27me3 across the broad domains typically observed at Polycomb targets, but can also contribute to the maintenance of these domains through cell division as the mark is diluted by the deposition of new nucleosomes after DNA replication.

Suppressor of Zeste (SUZ12) interacts with the catalytic subunit EZH2 or EZH1

of PRC2 primarily through its C-terminal VEFS domain, an interaction that aids in the stabilisation of the latter and is essential for enzymatic activity. This same VEFS domain has also been shown to interact with EED (Jiao and Liu, 2015), while the N terminus of SUZ12 interacts with both the RbAp46/48 subunit and several accessory factors involved in PRC2 recruitment and regulation. In addition, SUZ12 has been shown to be involved in the ability of PRC2 to sense the local chromatin state (see below, Recruitment of PRC2). The stimulation of the histone methyltransferase (HMT) activity of PRC2 in the presence of densely packed nucleosomes has been shown to involve SUZ12 and its interaction with a portion of histone H3 (Yuan et al., 2012). Conversely SUZ12 and its VEFS domain have also been suggested to be involved in the mechanism underlying the inhibition of H3K27 methylation by the pre-existence of H3K4me3 on the same histone tail (Schmitges et al., 2011), although the details of this mechanism remain poorly understood. SUZ12 may also be involved in the regulation of PRC2 recruitment as it has been shown to bind to CpG islands independently of the other core subunits (Højfeldt et al., 2018), but also binds to nascent RNA which was found to be antagonistic to PRC2 binding to chromatin (Beltran et al., 2016).

The association of the Retinoblastoma-binding protein 46 and 48 (RbAp46/48) subunit with PRC2 requires the presence of SUZ12, suggesting it only interacts with this subunit (Pasini et al., 2004). This subunit is known to interact with both histone H3 and H4 via its WD40 domain (Schmitges et al., 2011), possibly contributing to the stabilisation of PRC2 binding to nucleosomes. Its interaction with H3 has been shown to be prevented in the presence of H3K4me3, further hinting at the mechanism of inhibition of PRC2 activity by this mark (Schmitges et al., 2011).

Catalytic subunits Enhancer of zeste (EZH1/2)

By itself the EZH2 subunit assumes a catalytically inactive conformation and is unable to methylate H3K27 unless it forms a complex with at least EED and SUZ12. These interactions with EED and SUZ12 led to a change of state of the SET domain from an auto-inhibitory conformation to an active conformation in which the substrate binding site becomes accessible and the pocket which accepts the methyl donor SAM is stabilised (Jiao and Liu, 2015; Wu et al., 2013; Antonysamy et al., 2013; Justin et al., 2016).

All of these elements contribute to enabling the HMT activity of EZH2. This

activity can be further stimulated by the binding of H3K27me3 by the EED subunit, or inhibited by signals thought to be conferred by SUZ12 and RbAp46/48 in the presence of H3K4me3, as described above. The methyltransferase activity of PRC2 has been shown to be favoured by a more densely packed nucleosomal substrate (Yuan et al., 2012; Jiao and Liu, 2015). This is thought to be mediated at least in part by the SUZ12 subunit interacting with the H3 tail.

The enzymatically active SET methyltransferase domain of EZH2 itself has been extensively characterised and the crystal structure solved (Wu et al., 2013; Antonysamy et al., 2013; Jiao and Liu, 2015). Essential amino acids have been identified within or in proximity to the lysine binding pocket of the SET domain, several of which have been found to be mutated in cancerous cells leading to gain of function (Kipp et al., 2013; Ott et al., 2014; Wu et al., 2013; Antonysamy et al., 2013). This gain of function was found to be due to a shift in substrate preference of EZH2 to H3K27me2 leading to hypermethylation of H3K27 in these cells.

Additionally loss of function mutations have also been associated with malignancy, some of which include point mutations in the SET domain of EZH2 leading to abrogation of its catalytic activity while maintaining the integrity of the complex (Ernst et al., 2010; Jankowska et al., 2011). Recent studies have also achieved loss of function of EZH2 by the point mutation of a highly conserved tyrosine in physical proximity to the lysine pocket and the S-adenosyl-L-methionine, previously suggested to be essential to catalytic activity (Wu et al., 2013; Antonysamy et al., 2013; Lavarone et al., 2019; Wang et al., 2019).

The catalytic subunit of PRC2 exists in mammals as 2 paralogues named EZH1 and EZH2, which can each assemble with the other core subunits to form EZH1-PRC2 or EZH2-PRC2. The degree of conservation of these two proteins is shown in Fig. 1.1. These 2 catalytic paralogues of PRC2, although similar in sequence, are different from each other in several ways. Although both subunits are able to deposit trimethylation on H3K27, EZH2 displays a much higher HMT activity towards nucleosomes than EZH1 (Margueron et al., 2008). Additionally, a difference in the SRM domain of EZH1 contributes to a much reduced response in this subunit to the allosteric activation conferred by EED upon binding of H3K27me3, as compared to EZH2 (Lee et al., 2018). EZH1 and 2 also interact differently with chromatin. EZH1 has a higher affinity for chromatin than EZH2 (Son et al., 2013; Lee et al., 2018) due to nucleosome binding domains in the former that are poorly conserved in the latter. Electron

microscopy has also revealed a chromatin compaction activity of EZH1 that EZH2 does not possess (Margueron et al., 2008), although whether this ability is a direct consequence of the higher nucleosome binding affinity of EZH1 is not yet known. There is a suggestion that EZH1's higher affinity for chromatin may also contribute to the recruitment of EZH2 (Son et al., 2013). It has been shown that the binding of EZH2 to mononucleosomes *in vitro* is greatly enhanced by the addition of EZH1 and heterodimers of EZH1-PRC2 and EZH2-PRC2 have been identified *in vitro* and *in vivo* (Son et al., 2013; Davidovich et al., 2014). Lastly, the expression patterns displayed by EZH1 and EZH2 also differ from one another. While EZH2 expression tracks with highly proliferating cells (Bracken, 2003) and thus is the main subunit associated with PRC2 in ESCs, EZH1 is expressed more ubiquitously. As cells differentiate and slow rates of proliferation, EZH2 expression levels decrease whereas EZH1 is expressed in both adult and undifferentiated cell types, dividing and nondividing (Laible et al., 1997; Margueron et al., 2008). These differences between the 2 paralogues have given rise to a hypothesis of specialisation of function of these subunits wherein the more highly active EZH2 is required for the maintenance of H3K27me3 levels in actively proliferating cells, while EZH1 is sufficient to maintain these already established domains in more differentiated cells which are dividing much more slowly.

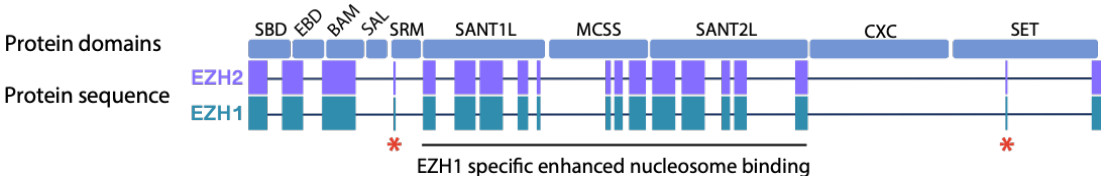


Figure 1.1 Comparison of EZH1 and EZH2 protein domains. Adapted from (Yu et al., 2019). Regions with the lowest sequence conservation between the two paralogues are highlighted in purple and green. Single residue differences in the SRM and SET domains that lead to the reduced activity of EZH1 are indicated with an asterisk.

1.2.2 Accessory subunits of PRC2

Several additional proteins have been found to interact with PRC2, diversifying the subtypes of the complex. These proteins are mostly involved in stabilising the complex on chromatin and enhancing or moderating its catalytic activity, and although many of them are involved in the localisation of PRC2 to its target sites,

the model for the exact mechanism of PRC2 recruitment remains incomplete. These accessory factors are not able to all interact with PRC2 at the same time and are in fact, in some cases, mutually exclusive. This has led to the further sub-classification of PRC2 into two types: PRC2.1 and PRC2.2 (Hauri et al., 2016; Holoch and Margueron, 2017). A schematic representation of these two distinct subtypes is shown in Fig. 1.2.

PRC2.1

PRC2.1 associates with a group of proteins referred to as PCL proteins. The *Drosophila* Polycomblike protein has 3 mammalian orthologues, PCL1,2, and 3, which area also named PHF1, MTF2 and PHF19, respectively. These proteins are all able to interact with SUZ12, however this binding occurs in a mutually exclusive manner and is also in competition with AEBP2 (Hauri et al., 2016; Grijzenhout et al., 2016). The Tudor domain and PHD fingers contained in all PCLs suggests a histone binding activity of these proteins that led to speculation that they could be involved in recruitment and chromatin binding of PRC2. For PCL2 an aspect of this function has been shown in its ability to facilitate recruitment of PCR2 to the inactive X chromosome (Casanova et al., 2011). This subunit has also been shown to contribute to the recruitment of PRC2 to CpG islands (Li et al., 2017; Perino et al., 2018).

Additionally, all 3 PCLs are able to recognise H3K36me3 through their N-terminal Tudor domain (Ballaré et al., 2012; Brien et al., 2012; Musselman et al., 2012), a mark associated with active genes which has been shown to be mutually exclusive with H3K27me3 on the same histone tail (Schmitges et al., 2011). The binding of an active mark by factors associated with a repressive complex such as PRC2 may seem contradictory, however it has been shown that for PCL1 and PCL3, this may serve as a way to target Polycomb repression to active genes (Cai et al., 2013; Brien et al., 2012). PCL3 specifically has been shown to enable the recruitment of PRC2 to these H3K36me3 associated genes while also recruiting NO66, a H3K36 demethylase, which would allow for clearing of the active mark and deposition of H3K27me3. Finally, PCL1 has been shown to be important for efficient tri-methylation of H3K27, as was also found to be the case for the *Drosophila* Pcl (Sarma et al., 2008; Nekrasov et al., 2007).

Recent studies have revealed the presence of an additional accessory factor of PRC2.1 whose function remains to be fully elucidated. EPOP or C17orf96 is

mostly unstructured and was found to interact with PRC2 via its C terminus (Liefke and Shi, 2015). The N terminus of this protein interacts with elongin BC (ELOBC) and it is thought that EPOP may function as a mediator between PRC2 and ELOBC, recruiting the latter to Polycomb genes in order to promote their transcription (Beringer et al., 2016; Liefke et al., 2016). It has been theorised that this may be a way to attenuate PRC2 activity and by maintaining low levels of transcription, allow for rapid activation in the event of the removal of Polycomb silencing. Another recently discovered interactor with PRC2.1 is Pali1/2. Found only in vertebrates, this protein was shown to stimulate the activity of PRC2 *in vitro* and associates with co-repressors in cells (Conway et al., 2018).

PRC2.2

Jumonji AT-rich interactive domain 2 (JARID2) is a member of the JMJC histone demethylases which are active towards methylated lysines. However, JARID2 lacks essential residues in the catalytic domain rendering it enzymatically inactive (Cloos et al., 2008). It has been shown by multiple studies to interact with PRC2 (Peng et al., 2009; Li et al., 2010; Pasini et al., 2010; Shen et al., 2009) and more recently this has been shown to involve its binding to the N-terminal ZnB-Zn domain of SUZ12 (Chen et al., 2018). JARID2 has been shown *in vitro* to stimulate the activity of PRC2 (Son et al., 2013), and its knockdown *in vivo* has been shown to lead to a reduction in PRC2 binding (Li et al., 2010; Pasini et al., 2010; Peng et al., 2009; Shen et al., 2009). Although the C terminus of JARID2 does have some DNA binding properties and has a preference for GC-rich sequences similar to those found at PRC2 occupied sites (Li et al., 2010), it is not yet known whether this activity contributes to PRC2 recruitment *in vivo*. JARID2 is also known to be involved in mediating feedback between PRC1 and PRC2, as it is able to bind the H2AK119ub mark placed by PRC1 through its N-terminal ubiquitin-interacting motif, which may also facilitate PRC2 recruitment (Kalb et al., 2014; Cooper et al., 2016). JARID2 has also been shown more recently to contribute to the stimulation of PRC2 activity. A residue in the N terminus of the protein, K116, was found to be a target for methylation by EZH2 (Sanulli et al., 2015). This lysine, once methylated, is able to stimulate the activity of PRC2 through its binding to the WD40 domain of EED, essentially mimicking the binding of a methylated H3K27 and possibly aiding in the establishment of H3K27me3 domains at previously unmethylated sites (Sanulli et al., 2015; Kasinath et al., 2018). JARID2 does

appear to have a preference for EZH2-PRC2 over EZH1-PRC2 *in vivo* (Pasini et al., 2010), and indeed its expression pattern closely mirrors that of EZH2, being highly expressed in undifferentiated, actively dividing cells and decreasing through differentiation (Mikkelsen et al., 2007; Walker et al., 2010). This reduced association with EZH1 correlates with this subunit's increased innate stability on chromatin and relative lack of response to stimulation by EED.

A protein with functions complimentary to JARID2 is the zinc-finger protein Adipocyte enhancer-binding protein 2 (AEBP2). This accessory factor was found to increase the stability of the complex when associated with PRC2 (Cao et al., 2002; Ciferri et al., 2012). This interaction occurs primarily through contacts between the C terminus of AEBP2 and the C2 domain of SUZ12 (Chen et al., 2018) but also involves other members of PRC2 (Cao and Zhang, 2004; Ciferri et al., 2012). Although AEBP2 possesses DNA binding activity through its zinc-finger domain, there is little sequence specificity which, as with JARID2, does not provide strong evidence for the contribution of this activity to site-specific PRC2 recruitment but might aid in increasing general chromatin affinity of the complex (Kim et al., 2009). AEBP2 has also been shown to bind to RbAP46/48, mimicking its interaction with the unmodified H3 tail. This interaction further contributes to stability of the complex, which in turn enhances stimulation by JARID2 (Son et al., 2013; Kasinath et al., 2018).

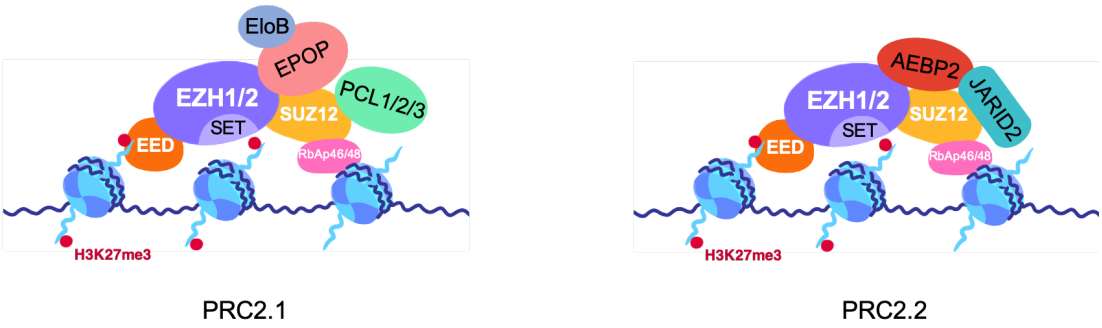


Figure 1.2 Schematic representation of the PRC2.1 and PRC2.2 complexes. The association of the PRC2 core complex (EZH1/2, EED, SUZ12 and RBAP46/48) with its accessory subunits are divided into two subcomplexes, PRC2.1 and PRC2.2.

1.2.3 Catalytic products of PRC2 and their functions

Although the principle histone modification associated with PRC2 is H3K27me3, it is also responsible for the deposition of mono and di-methylation at this residue. Each of these 3 methylation states of H3K27 has a distinct profile of distribution and abundance throughout the genome. In ESCs around 70% of all histone H3 carries H3K27me2 while H3K27me1 and H3K27me3 occupy a further 10% and 5% of H3 respectively (Peters et al., 2003; Ferrari et al., 2014). Perhaps the most poorly understood of the 3 methylation states of H3K27, H3K27me1 is paradoxically enriched in the gene bodies of actively transcribed genes and correlates with H3K36me3 (Ferrari et al., 2014). In some studies it has been found to persist in PRC2 deficient cells which gave rise to speculation of an additional pathway of deposition of this mark (Schoeftner et al., 2006; Pasini et al., 2007), although it has since been shown is that it relies solely on PRC2 (Ferrari et al., 2014). H3K27me2 is the most abundant of all forms of H3K27 methylation, is found in much broader regions and is associated with a more transcriptionally inactive state (Ferrari et al., 2014). This mark also covers the flanking regions surrounding H3K27me3 enriched domains and its enrichment accumulates at these sites before the catalysis of trimethylation (Oksuz et al., 2018). The enrichment of H3K27me3 broadly mirrors that of PRC2 being found mostly around the TSS of unmethylated CGI promoters, although in some cases can cover broader domains such as at the Hox gene clusters where it can spread over more than 100 kb (Boyer et al., 2006; Zhao et al., 2007). As mentioned above, PRC2 is able to bind to this histone modification via the EED subunit, in a manner that is thought to aid in the propagation of H3K27me3 (Margueron et al., 2009; Oksuz et al., 2018; Poepsel et al., 2018). This mark also acts as one mode of recruitment of PRC1 via the CBX subunit, as will be further explained below. The depostion of H3K27me2 does not require stable binding of PRC2 to chromatin and this mark is deposited rapidly following DNA synthesis (Sneeringer et al., 2010). However, PRC2 has a higher activity towards unmethylated H3K27 than towards H3K27me2 and thus the deposition of H3K27me3 is much less efficient and mostly co-localised with the binding of PRC2 itself (Tie et al., 2009). This suggests that H3K27me3 is only able to be catalysed in regions where there are additional factors stabilising the binding of PRC2, and in this way the deposition of this mark can be more specific and fine-tuned as compared to the broad regions of mono and dimethylation of H3K27. This in turn also leads to a more site specific recruitment of PRC1 via its binding to H3K27me3. The

dimethylation of H3K27 appears to not play an important role in gene silencing, however it is possible that this mark may not exist solely as an intermediate to H3K27me3, but also as a way to inhibit the deposition of the active H3K27ac mark which is antagonistic to Polycomb gene repression (Trojer and Reinberg, 2007).

Additional non-histone targets of PRC2 methyltransferase activity have also been discovered. As previously mentioned the accessory protein JARID2 can be methylated on its lysine 116 (Sanulli et al., 2015). Other examples include GATA4, the methylation of which reduced the transcriptional activity of this protein (He et al., 2012), and STAT3, for which the activity is stimulated following methlyation (Kim et al., 2013). Recently, an additional non-histone target of PRC2 methylation was identified as Elongin A (Ardehali et al., 2017). The methylation of this transcription elongation factor is thought to contribute to the maintenance of a repressed state at a subset of Polycomb genes.

1.2.4 PRC1

The PRC1 complex is one of the main binders of H3K27me3 and is implicated in mediating a significant portion of its function. PRC1 encompasses a larger collection of subunits than have been identified for PRC2, and as such this complex can assemble in a number of different combinations leading to a diverse set of variants.

Core subunits and variant complexes

Although the PRC1 complex is subdivided into multiple variants, there are 2 core subunits that are common to all versions of the PRC1 complex. The first of these is RING1A/B, and the second is PCGF, of which there are 6 paralogues in mammals. Together these two subunits, function as an E3 ubiquitin ligase and are responsible for the ubiquitylation of lysine 119 on histone H2A (H2AK119ub) (Wang et al., 2004; McGinty et al., 2014). The PRC1 complexes that assemble around this core heterodimer are defined as canonical or non-canonical based on the subset of additional proteins that they contain, as well as which of the PCGF subunits is included (Gao et al., 2012). The canonical complex contains one of the 5 CBX proteins CBX2,4,6,7, or 8, all of which contain a chromodomain which is able to bind to H3K27me3, as well as a PHC subunit

which contributes to gene repression by oligomerisation of its SAM domain leading to chromatin compaction (Fischle, 2003; Bernstein et al., 2006b; Isono et al., 2013). This compaction activity is discussed in more detail in section 1.5. The non-canonical PRC1 contains the RYBP/YAF2 subunit which binds to the core heterodimer in a mutually exclusive manner with the chromobox (CBX) proteins. The non-canonical complex can also include H3K36 demethylase KDM2B which is able to bind to CGIs via its zinc-finger CxxC domain (Farcas et al., 2012), as well as DCAF7 and WDR5 both of which contain WD40 domains and act as scaffold proteins (Hauri et al., 2016). There are several different proteins that can additionally associate with non-canonical PRC1, further diversifying the subsets of this variant and fine-tuning its activity (Gao et al., 2012). These variants often depend on the subunits expressed in different cell types (O'Loghlen et al., 2012). Although both canonical and non-canonical PRC1 are able to ubiquitylate H2AK119, it has been shown both *in vivo* and *in vitro* that of the two, the non-canonical complex has the higher catalytic activity (Blackledge et al., 2014; Gao et al., 2012). This mark has been suggested to mediate gene repression in part by inhibiting transcriptional elongation (Zhou et al., 2008; Stock et al., 2007). However, H2AK119ub is not essential for Polycomb mediated silencing as its depletion does not lead to upregulation of target genes in *Drosophila* (Pengelly et al., 2015), and it is thought that the chromatin compaction activity of PRC1 plays a more substantial role in repression (Isono et al., 2013; Lau et al., 2017).

1.3 Polycomb target genes and recruitment

As previously mentioned, PcG proteins were first discovered as chromatin binding proteins that were responsible for the timely regulation of Hox genes during the embryonic development of *Drosophila* (Lewis, 1978). Since their discovery this association with the regulation of developmental genes has persisted and been shown to be critical in mammalian development (O'Carroll et al., 2001; Pasini et al., 2004; Bernstein et al., 2006a). PcG proteins occupy a highly conserved set of genes (Schuettengruber et al., 2007) which in mammals is highly enriched for transcription factors involved in the regulation of developmental pathways, as well as other factors involved in cell fate decisions and differentiation (Boyer et al., 2006; Bracken, 2006; Lee et al., 2006). A well as their role as regulators of developmental genes, PcG protiens have additionally been found to have functions in X chromosome inactivation and silencing of imprinted genes (Zhao et al., 2008;

da Rocha et al., 2014; Inoue et al., 2017).

In line with their binding to developmental genes, many Polycomb mutants cause severe developmental abnormalities. The mutants generated in *Drosophila* led to striking defects in body patterning and limb development (Lewis, 1978). In ESCs, depletion of PcG proteins has little effect on gene expression, nor on the self-renewal capacity or viability of the cells (Chamberlain et al., 2008; Shen et al., 2008; Riising et al., 2014). These deficient cells are however unable to differentiate correctly in many cases (Chamberlain et al., 2008; Shen et al., 2008; Pasini et al., 2007). In mice, null mutations of PcG proteins lead to defects early in embryonic development. Knockout mutants of core proteins of the PRC2 complex result in developmental defects that become evident at the post implantation stage and are usually embryonic lethal around the stage of gastrulation (Faust et al., 1995; O'Carroll et al., 2001; Pasini et al., 2004) with the knockout of RING1B leading to a similar lethality at a slightly later stage (Voncken et al., 2003). Null mutants of PRC2 accessory factors, and all other PRC1 subunits however, have mostly been found to produce viable mice, although with varying degrees of abnormalities and growth defects (Coré et al., 1997; Motoyama et al., 1997; Takihara et al., 1997; del Mar Lorente et al., 2000; Walker et al., 2010).

1.3.1 Role of CpG islands in Polycomb recruitment

One of the key features of Polycomb bound genes is the presence of unmethylated CpG islands in proximity to the transcriptional start sites (TSS) at promoters (Boyer et al., 2006; Tanay et al., 2007; Mendenhall et al., 2010). It has even been shown that the introduction of ectopic, unmethylated CG rich DNA was sufficient to induce recruitment of PcG proteins in ESCs (Mendenhall et al., 2010; Jermann et al., 2014; Riising et al., 2014; Wachter et al., 2014). This discovery led to the hypothesis that CGIs may serve as the mammalian equivalent of the Polycomb response elements (PREs) found in *Drosophila* which serve to recruit PRC2 in this organism. However there is little evidence that direct binding to CGIs is responsible for all Polycomb recruitment. Both JARID2 and PCL2 possess low affinity DNA-binding activity with a preference for CG-rich sequences, and it has recently been shown that a knockout of both these proteins leads to a massive loss of PRC2 binding to chromatin (Oksuz et al., 2018). However, there was evidence that PRC2 could still access some of its targets in this context, hinting at additional factors involved in its recruitment (Oksuz et al., 2018). One

such alternative recruitment method to CG-rich sequences is through the PRC1 associated protein KDM2B which can be included as part of the noncanonical complex (Lagarou et al., 2008; Farcas et al., 2012). Although this protein is a demethylase of H3K36, it also possesses a zinc-finger CxxC domain which binds to CGIs allowing the recruitment of PRC1 to these sites (He et al., 2013). Despite this activity and its near ubiquitous binding of CGIs however, KDM2B has only been shown to be essential for the recruitment of Polycomb to around one third of all mammalian CGIs (Bernstein et al., 2006b; Ku et al., 2008) but could act as the starting point for the PRC1 dependent recruitment of PRC2 in some instances. A few additional DNA binding factors such as REST and SNAIL have been found to interact with PcG proteins and contribute to recruitment, however each of these have only proven to be involved at specific loci and not essential for Polycomb recruitment as a whole (Herranz et al., 2008; Dietrich et al., 2012; Yu et al., 2012).

As is suggested by the preference of both PRC1 and PRC2 for unmethylated CGIs, DNA methylation is an important factor in defining the borders of Polycomb domains. Hypomethylation induced by the depletion of the Dnmts has been shown to lead to spreading of H3K27me3 into regions not normally occupied by this mark (Brinkman et al., 2012). This antagonism of Polycomb activity by DNA methylation appears to lie in its ability to obstruct the binding of PRC2 to chromatin (Bartke et al., 2010). Methylated DNA is not however the only factor restricting the spreading of Polycomb domains, as hypomethylation of the genome does not lead to the deposition of H3K27me3 at all newly demethylated sites (Brinkman et al., 2012). Indeed, there is some evidence that PRC2 may not even be directly affected by DNA methylation. In *in vitro* assays of PRC2 methyltransferase activity methylated DNA appeared to have no effect on catalysis (Cooper et al., 2014). Furthermore recruitment of PRC2 to a methylated region was achieved by artificial deposition of H2AK119ub, as was methylation of H3K27 (Cooper et al., 2014), indicating that methylated DNA is not as inimicable to PRC2 binding as was suggested by their lack of overlap on chromatin.

1.3.2 Role of RNA in PRC2 activity

The interaction of Polycomb with RNA has become apparent in recent years and although the role of lncRNAs in the recruitment of PcGs to specific loci has been established, expansion of this model of recruitment to include other RNAs in a more general Polycomb targeting strategy remains elusive. The first instance of

lncRNA involvement in PRC2 recruitment was shown with the interaction of *Xist* (X inactive specific transcript) with PRC2 in its localisation to the inactive X chromosome (Silva et al., 2003). This process requires the presence of JARID2 and may not be a direct interaction but rather may involve an additional *Xist* binding protein known as SHARP (da Rocha et al., 2014; Cerase et al., 2014; McHugh et al., 2015). A further example of Polycomb recruitment by lncRNA involves *HOTAIR*. This transcript originates from the Hox C locus and its direct interaction with PRC2 serves to recruit the complex *in trans* to the HoxD locus (Rinn et al., 2007).

In addition to its interactions with specific lncRNAs, PRC2 has been show in recent years to interact with RNA in general with low sequence specificity (Davidovich et al., 2015; Beltran et al., 2016) and indeed is even able to bind to L-RNA, the enantiomer of naturally occurring RNA molecules (Deckard and Sczepanski, 2018). Despite the evidence of these many interactions, their involvement in PRC2 function is still unclear and appears to vary based on the genomic context. In particular, the association of PRC2 with nascent transcripts at active genes has been a point of debate (Kaneko et al., 2013; Kaneko et al., 2014). Two schools of thought exist on this subject, one argues that the binding to nascent RNAs sequesters PRC2 away from chromatin at active genes in order to inhibit any repressive effects of PRC2 on transcription (Beltran et al., 2016). The second theory views this binding of nascent transcripts as a method of localisation of PRC2 in proximity to the promoter region of active genes for timely recruitment to chromatin in the event of transcriptional repression of these genes (Kaneko et al., 2014). In the case of transcriptionally repressed genes it is thought that the binding of PRC2 to short, non-elongating RNAs is able to stabilise the association of the complex and repress any further aberrant initiation of transcription (Kanhere et al., 2010). It should be noted that RNA has been shown to inhibit the catalytic activity of PRC2 in vitro (Cifuentes-Rojas et al., 2014; Kaneko et al., 2014) which further hints at its role being to attenuate the catalytic activity of PRC2 towards its chromatin substrate.

It has recently been shown that the formation of a DNA-RNA hybrid structure called R-loops over a subset of Polycomb genes contributes to both gene repression at these sites, and the recruitment of PcG proteins (Skourti-Stathaki et al., 2019). R-loops are formed by the interaction of nascent RNA with the template DNA strand which leaves the opposite DNA strand single-stranded over this region and have been shown to mediate transcriptional repression (Skourti-Stathaki et al.,

2011). The resolution of these R-loops in mESCs led to a decrease in RING1B recruitment and derepression of genes that appeared to be independent of PRC2 activity (Skourti-Stathaki et al., 2019).

1.3.3 Chromatin features involved in Polycomb occupancy

Mutual recruitment of Polycomb group proteins

PRC1 and PRC2 co-occupy many of their target genes, which in turn leads to the co-occurrence of the H3K27me3 and H2AK119ub marks on chromatin. There is an interaction of these 2 complexes based on the mutual recognition of the histone mark placed by the other complex, for PRC1 through the binding of the CBX subunit to H3K27me3 and for PRC2 through the interaction of JARID2 and AEBP2 with H2AK119ub. This first interaction of CBX containing PRC1 with H3K27me3 has been shown to increase its affinity for chromatin (Fischle, 2003). This is known as the hierarchical model of Polycomb recruitment and involves initial recruitment of PRC2 which then deposits H3K27me3 which in turn leads to the recruitment of a canonical PRC1 through its CBX subunit, and subsequent ubiquitilation of H2AK119 (Wang et al., 2004). In this model PRC2 serves as the initial marker of genes to be silenced and mediates the recruitment of PRC1 which is more directly responsible for gene repression by the deposition of H2AK119ub and the formation of compact Polycomb bodies.

Although this method of PRC1 recruitment makes sense for the canonical form of the complex, this H3K27me3 dependent model does not work with the non-canonical PRC1 complex which does not possess the H2K27me3 binding CBX subunit. Recruitment of PRC1 to chromatin independently of PRC2 has since been observed (Tavares et al., 2012). An alternative model for recruitment of PcG proteins was put forth following a study involving tethering PRC1 to a region not normally occupied by PcG proteins. This tethering of PRC1 and resulting H2AK119ub were found to lead to the recruitment of PRC2 and deposition of H3K27me3 at this locus (Blackledge et al., 2014; Cooper et al., 2014). The recruitment of PRC2 to PRC1 bound loci is thought to involve the interaction of a JARID2 and AEBP2 associated PRC2 with the H2AK119ub catalysed by PRC1 (Kalb et al., 2014) and is thought to predominantly involve the more catalytically active noncanonical form of PRC1 (Blackledge et al., 2014; Tavares et al., 2012). In this model gene repression is still thought to be mainly mediated

by canonical PRC1 which can be recruited via the hierarchical mechanism once PRC2 has been recruited by the noncanonical PRC1.

Influence of additional histone modifications and role of transcriptional state

As mentioned above, both PRC1 and PRC2 are able to sense the histone mark placed by the other complex and this aids in their recruitment. This mutual recruitment of the 2 complexes gives rise to the co-occupation of most Polycomb bound genes by both complexes, however this is not the case at all of the target sites (Ku et al., 2008). The methylation of H3K36 may also contribute to the recruitment of PRC2 in some cases both through its interaction with the tudor domains of the PCL proteins (as mentioned above), facilitating the establishment of Polycomb silenced domains at previously active genes (Ballaré et al., 2012). There is also a suggestion that H3K9me3 may influence PRC2 occupancy on chromatin as this mark can be bound by EED (Margueron et al., 2009) and the enzymes that place it can interact with the PRC2 complex (Mozzetta et al., 2014) although this histone modification is associated with constitutive heterochromatin rather than the facultative heterochromatin of Polycomb domains.

Several elements of active gene transcription have been shown to be antagonistic to Polycomb activity. Indeed transcription inhibits the deposition of H3K27me3 in mESCS (Jermann et al., 2014). Conversely, global inhibition of transcription has been shown to be sufficient to recruit PRC2 to many newly inactive sites (Riising et al., 2014). As previously discussed, the nascent transcripts from active genes are thought to sequester PRC2 from chromatin and histone modifications associated with active marks also contribute to obstructing Polycomb activity at these sites. Marks such as H3K4me3 and H3K36me3 have been shown to exclude the deposition of H3K27me3 on the same histone tail, potentially through a mechanism involving SUZ12 and RbAp46/48 (Schmitges et al., 2011). In the case of H3K27ac, another mark associated with active genes, its presence physically excludes the deposition of methylation on H3K27. However all of these active marks can coexist in the same nucleosome as H3K27me3 if residing on separate H3 tails (Voigt et al., 2012). This finding, along with the fact that complexes that place active marks such as the SET1A/B and MLL1/2 complexes have a similar preference for unmethylated CGIs as Polycomb, goes some way to explaining why active H3K4me3 and repressive H3K27me3 have been found to co-occur in many

instances, despite there apparent opposite functions.

1.4 Bivalency

1.4.1 Discovery in ESCs

It was discovered that in ESCs virtually all Polycomb genes display a co-localisation of both active and repressive marks. These sites were first discovered in 2006 by mapping H3K4 and H3K27 methylation across highly conserved non-coding elements such as promoters and enhancers by chromatin immunoprecipitation (ChIP) and microarray analysis (Bernstein et al., 2006a). These experiments found that at the promoters of some genes there is an overlap of the repressive H3K27me3 and the active H3K4me3 marks. These loci, which are mainly found at developmental genes in ESCs, were dubbed bivalent domains. These studies were confirmed by genome-wide ChIP and next generation sequencing (ChIP-seq) studies in ESCs but also, to lesser extent, other cell types (Mikkelsen et al., 2007; Pan et al., 2007; Mohn et al., 2008). A bivalent state was later also found to exist at enhancers of some developmental genes, again in ESCs. In the case of enhancers the bivalent or poised state is defined by a co-localisation of H3K27me3 and H3K4me1 (Creyghton et al., 2010; Rada-Iglesias et al., 2011).

1.4.2 Mechanism

As mentioned above, H3K4me3 can be placed both by the SET1A/B complexes and the MLL1/2 complexes, although it is thought that at bivalent promoters this mark is preferentially catalysed by the MLL2 complex (Clouaire et al., 2012; Denissov et al., 2014; Hu et al., 2013). Both the SET1A/B and MLL1/2 proteins belong to the Trithorax family and either themselves (MLL1/2) or indirectly (CFP1 subunit of SET1 complexes) contain CxxC domains that allow them to recognise unmethylated CpG islands (Birke, 2002; Bach et al., 2009), found at high density at the vast majority of mammalian promoters and frequently targeted by Polycomb.

The bivalent state of chromatin at Polycomb bound enhancers is mediated by a slightly different set of complexes. H3K4me1 is placed by the MLL3/4 complexes

and is mostly enriched at active enhancers (Herz et al., 2012). This mark also has a broad distribution across gene bodies however, this is likely a by-product of SET1A/B-dependent catalysis of H3K4me3 at promoters. Despite also belonging to the Trithorax family, MLL3/4 do not possess a CxxC domain, which may explain why they are not enriched at promoters. However, it remains unclear how MLL3/4 are specifically targeted to enhancers. Presence of H3K4me1 has been shown to often precede and indeed stimulate, at least in *Drosophila*, deposition of the H3K27ac mark by the CBP/p300 co-activator proteins and subsequent activation of enhancers (Tie et al., 2014). CBP/p300 has been found to be in physical contact with the H3K27 demethylase UTX. A subset of MLL3/4 complexes are known to contain UTX, suggesting a process by which MLL3/4 in mono-methylating H3K4 could allow the removal of methylation at H3K27 by UTX, after which it is then accessible to acetylation by CBP/p300 (Herz et al., 2012).

As has been discussed previously, the presence of active histone modifications is known to be antagonistic to PRC2 activity, however the presence of these active marks on one H3 histone within the nucleosome does not preclude the deposition of H3K27me3 on the other H3. Indeed it was shown that nucleosomes at bivalent domains are largely in an asymmetric conformation, with one copy of H3 carrying the active mark and the other the repressive one (Voigt et al., 2012; Shema et al., 2016). Nucleosomes at bivalent domains could also be symmetrically modified with active or repressive PTMs and bivalency conferred by positioning of active and repressive nucleosomes adjacent to one another. Recent genome wide reChIP studies have further confirmed this co-occurrence of H3K4me3 and H3K27me3 on the same nucleosomes (Kinkley et al., 2016; Sen et al., 2016; Shema et al., 2016; Weiner et al., 2016; Mas et al., 2018).

1.4.3 Hypotheses of function

The purpose of bivalent domains is thought to be to keep genes in a poised state in ES cells, keeping transcription largely repressed but ready to be activated or fully repressed rapidly upon cell differentiation, simply by the removal of one or the other of the opposing PTMs. Indeed, the bivalency at these loci is mostly resolved upon cell differentiation as H3K27me3 is lost and genes are expressed or H3K4me1/3 is lost and genes are silenced (Mikkelsen et al., 2007; Pan et al., 2007). The subset of genes that become silenced or expressed by the

resolution of these bivalent domains depends largely on which cell type results from the differentiation process. It appears that the bivalent state is at least dispensable for viability and the maintenance of self renewal in ESCs as these are largely unaffected by the loss of bivalent chromatin in EED-/- or SUZ12-/- cells (Pasini et al., 2007; Chamberlain et al., 2008; Shen et al., 2008). One function that H3K4me3 may have at these domains is to keep the Polycomb genes in a more permissive state than they would otherwise be. In mESCs, depletion of MLL2 was found to lead to greater compaction and reduced transcription of many bivalent genes following an increase in binding of PcG proteins (Mas et al., 2018). Depletion of MLL2 in these cells also leads to impaired differentiation (Lubitz et al., 2007; Denissov et al., 2014) and in mice is lethal by E11.5, slightly later than seen in many Polycomb knockouts (Glaser et al., 2006).

Although Polycomb genes are not actively transcribed, there is still evidence of RNA polymerase II bound at their promoters. This Pol II is not in an active state, as demonstrated by its phosphorylation signature, however there are still detectable levels of RNA transcribed from these genes (Brookes et al., 2012; Stock et al., 2007). This low level of transcription only generates short, immature RNAs as there is no elongation following initiation. Nevertheless, this is a further element arguing for the function of Polycomb regulation being to maintain a poised state of gene expression rather than a fully silenced state. Another characteristic of Polycomb genes is an enrichment of the histone variant H2Az (Creyghton et al., 2008). This histone is mostly associated with active genes and has also been shown to be mutually exclusive with DNA methylation, which is consistent with the state observed at bivalent genes.

Although we have some insights into how these different PTMs such as H3K27me3 and H3K4me3 work individually, there is little known about how they function together, particularly in combinations such as those found at bivalent domains. It remains unclear how this bivalency is established and maintained in ESCs, particularly given that the enzymes that place the opposing marks are mutually antagonistic. It is also unclear whether these complexes have non-catalytic roles in addition to their histone modifying activity, for instance by occupying binding sites they could block access to other factors, or they could be involved in nuclear organisation, mediating interactions between different sites.

The exact function of these bivalent domains in regulating gene expression and cell differentiation also remains to be explicitly defined. Poised enhancers in particular have not been extensively studied and much about their regulation,

interactions and specific function in determining cell fate are still unknown.

1.5 Involvement of Polycomb in chromatin organisation

A significant branch of epigenetic regulation is the modulation of chromatin state into more "open" or "closed" conformations which are thought to affect transcription by controlling the accessibility of DNA to binding factors. It has been established for some years that Polycomb domains are in a more condensed state than that of actively transcribed genes (Shao et al., 1999; Francis et al., 2004).

Of the PcG proteins, those most associated so far with changes to chromatin structure are members of the PRC1 complex. While the non-canonical form of PRC1 is responsible for the majority of H2AK119ub, the canonical form of the complex is able to influence chromatin structure, both on a local and a more distal scale. Polycomb domains are characterised by a more condensed chromatin state, nucleosome density is increased at the promoters of Polycomb bound genes with a slower turnover of histones. However, the incorporation of the more mobile histone variant H2Az in these nucleosome may counteract, to a certain extent, the decrease in accessibility that would be generated by the condensation of chromatin at theses sites. As PRC2 has been shown to be more active towards more densely packed nucleosomes (Yuan et al., 2012), this system may function as a way to enhance the Polycomb signature at these sites, while maintaining the more open state of chromatin typical of Polycomb domains, that is more permissive to transcription than constitutive heterochromatin. The mechanism by which this local condensation of chromatin occurs involves the CBX2 subunit of the canonical PRC1. The intrinsically disorganised region (IDR) of this protein has been shown to be essential for this compaction (Lau et al., 2017).

In regions where multiple Polycomb targets are localised in close proximity along the chromosome, such as at the Hox gene clusters, this chromatin compaction occurs on a much larger scale. Across these broad domains the PHC subunits become more important for the condensing activity of PRC1. The SAM domains of this subunit are able to oligomerise to bring into contact multiple Polycomb targets whilst also omitting any non-Polycomb genes that may exist in the same

region (Isono et al., 2013; Kundu et al., 2017). This activity has also been observed on a much larger scale, with chromatin looping between distal Polycomb domains, even between separate chromosomes. Many of these contacts have been shown to dissociate upon cell differentiation as the genes involved become expressed, and indeed in some studies the loss of condensing activity of PRC1 has been shown to lead to derepression of Polycomb targets. Whether on a local scale or involving much more distal contacts, this chromatin compaction activity of PRC1 has been shown to function in a manner that is independent of its catalytic activity (Eskeland et al., 2010). Indeed, depletion of H2AK119ub was found to have little effect on gene repression or on chromatin compaction in mESCs (Illingworth et al., 2015; Kundu et al., 2017).

While the majority of studies of Polycomb involvement in chromatin structure have focused on PRC1, there is also evidence of PRC2 influencing these processes. PRC2 has been associated with the formation of long-range chromatin contacts in multiple studies. The depletion of PRC2 has been shown to disrupt structures formed both in the clustering together of multiple Polycomb targeted regions, within a single chromosome and between chromosomes (Denholtz et al., 2013; Vieux-Rochas et al., 2015; Tiwari et al., 2008a), as well as those involving chromatin looping to bring into contact specific loci (Tiwari et al., 2008b; Joshi et al., 2015).

Although in general the chromatin contacts and compaction mediated by Polycomb have mostly been associated with repression of transcription, in some instances it has been suggested to function to mediate the contact between genes and regulatory elements in preparation for their activation upon cell differentiation. In ESCs, poised enhancers have been found to already be in contact with the promoters of some Polycomb bound genes, although in contrast to active enhancers, this contact does not trigger active transcription of these genes (Cruz-Molina et al., 2017; Kondo et al., 2014; Gentile et al., 2019). How this looping is established is not fully understood but it has been suggested that it may be in part mediated by PRC2 as the loss of this complex leads to the dissociation of these contacts. The looping of these enhancers and promoters, while not resulting in active transcription in undifferentiated cells was found to be essential for the expression of these genes at the correct time during development.

Although it is thought that many of the contacts that have been shown to be PRC2 dependent may only be so insofar as PRC2 is required for the recruitment of PRC1, which then is directly responsible for the chromatin compaction, it has

yet to be fully proven that PRC2 plays no role in this independently of PRC1. Furthermore, it has not been definitively proven whether the catalytic activity of PRC2 is a critical factor in the establishment of condensed chromatin.

1.6 Aims of the project

As discussed above, the PRC1 complex, although responsible for placing a histone modification, also has functions that are independent of this catalytic activity. A recent study involving the H3K4 methyltransferases MLL3/4 has shown that the catalytic activity of these complexes is not required for the maintenance of enhancer RNA transcription and Pol II occupancy of their target genes (Dorighi et al., 2017). However, a similar study on these complexes found that the loss of catalytic activity led to a destabilisation of the MLL4 protein (Jang et al., 2017). Given these findings concerning complexes involved in similar chromatin modifying activities, and the complexity of the PRC2 interactome, it seems reasonable to hypothesise that PRC2 may perform functions that do not rely entirely upon its catalysis of the H3K27me3 mark. In this project I aimed to further our knowledge on H3K27me3's role in the execution of PRC2 function. To achieve this, I generated knockout mutant mESCs that lead to a depletion of both the PRC2 complex and H3K27me3 as well as catalytically inactive PRC2 mutant mESCs in which only the H3K27me3 was lost. I compared these two PRC2 mutants in order to address the following points: First, how does the loss of catalytic activity of PRC2 affect its binding to Polycomb target sites and how does this influence H3K4me3 levels at bivalent genes? Second, how crucial is the H3K27me3 mark for the maintenance of 3D chromatin structures at specific Polycomb targets? Third, how does the loss of H3K27me3 alone affect gene expression as compared to the loss of both the PRC2 complex and the histone mark it places? Finally, I aimed to clarify whether the catalytic activity of PRC2 is essential for differentiation and if so at what stage.

Chapter 2

Materials and Methods

2.1 Cell culture

2.1.1 mESC culture conditions

E14 male Mouse embryonic stem cells (mESCs) were grown at 37°C in 5% CO_2 on 10 cm plates coated with 0.1% gelatin (Sigma) in ESC media: DMEM containing 4.5 g/l glucose (Gibco) with 15% fetal bovine serum (FBS, Life Technologies, South American, A3160802), 2 mM l-glutamine, 1 mM pyruvate, 1x MEM nonessential amino acids, 50 units/ml penicillin and 50 µg/ml streptomycin (all Gibco), 0.2 mM -mercaptoethanol (Sigma), and heterologously expressed recombinant LIF (made in-house and batch-tested for maintenance of self-renewal).

2.1.2 Cell passaging

mESCs were passaged every 2 days at a ratio of around 1:15. Media was removed and cells were briefly washed in around 5 ml of warm PBS (Gibco) before adding 2.5 ml 0.25% trypsin-EDTA (Gibco) and incubating cells at 37°C for 5 min. Trypsin was then quenched by adding 7.5 ml of ESC media and cells were gently resuspended using a 5 ml serological pipette (Sarstedt) before being centrifuged at 500 x g for 5 min at room temperature. The cell pellet was then resuspended in ESC media and plated at required density onto fresh gelatinised plates.

2.1.3 Cryo-preservation of cells

Cells that were to be frozen for storage in liquid nitrogen were cultured as described above and grown on the required number of 15 cm plates before being trypsinised for 5 mins at 37°C. Trypsin was quenched and resuspended as described above before cells were counted and centrifuged at 500xg for 5 mins at room temperature. The cell pellet was then resuspended in the appropriate volume of serum free freezing medium Bambanker (Lymphotec) for a density of 4 million cells per ml. Cells were then aliquoted into cryogenic vials (Corning) with 2 million cells per vial before being added to a freezing container (Nalgene) which was then placed in a -80°C freezer. The next day, they were then transferred to liquid nitrogen storage.

2.2 Generation of mutant cell lines: CRISPR design, transfection, and validation

The approach taken for the CRISPR/Cas9 genome editing followed the protocol laid out by the Zhang lab in 2013 (Ran et al., 2013).

2.2.1 gRNA and ultramer design

Optimal gRNA sequences were chosen from those suggested by the Zhang lab CRISPR design tool (crispr.mit.edu url now defunct). This tool uses an algorithm to score gRNAs for a given target sequence based on the likelihood of off-target binding (Hsu et al., 2013; Ran et al., 2013). Single stranded donor oligodeoxynucleotides (ssODN, ordered from IDT) of around 200 bp were designed to act as donors for homologous repair of targeted sites. These ssODNs contained the desired mutations for each genoytpe with additional point mutations in the gRNA hybridisation site and/or in the PAM to prevent cleavage of the donor and repeated cleavage events by Cas9 after repair of the target sequence. ssODNs were designed with at least 70 bp of homology flanking each side of the targeted site to enable efficient homologous repair. Single strand oligos for gRNA expression were annealed and phosphorylated before being ligated into linearised into pSpCas9(BB)-2A-GFP (pX458) Cas9 expression vector following the protocol described in Ran et al. (Ran et al., 2013). This vector enables the

expression of the *S. pyrogenes* Cas9 endonuclease as well as GFP, and contains the cloning backbone for the gRNA. Ligated plasmids were then used to transform chemically competent XL-10 gold strain *E.coli*. Individual colonies were picked and expanded in 5 ml LB medium (10 g/l tryptone, 5 g/l yeast extract, 10 g/l NaCl) for small-scale plasmid preparation (EZNA plasmid mini kit, Omega). sgRNA inserts in purified plasmids were sequenced using a primer annealing to the the U6 promoter and bigdye sequencing technology (Thermofisher)according to the manufacturer's instructions. Sequencing runs were carried out by Edinburgh Genomics. The detail of sgRNA sequence and ssODN design for each cell line can be found in chapters 3 and 5.

2.2.2 Generation, transfection, cell sorting and colony picking/Genotyping

Co-transfection of the sgRNA and Cas9 in pX458 and donor ssODN was performed using Optimem and Lipofectamine 2000 on low passage E14 ES cells.

400,000 cells were used per transfection and were grown in 6-well plates for 24 h post-transfection without antibiotics and for another 24 h with antibiotics. For FACS sorting, cells were trypsinised and passed through a 70 µm cell strainer. FACS sorting of GFP+ cells (co-expressed with Cas9 from pX458 as a 2A fusion) was performed and 10,000-15,000 cells were then plated onto a 15 cm plate. After 7-10 days, single colonies were picked by hand, broken up with trypsin, and plated in duplicate into two 96-well plates, one for genotyping and one for maintaining the cell lines, and grown for 2-3 days. DNA was extracted from one plate using the QuickExtract DNA extraction solution (Epicentre) and an initial genotyping PCR was performed using primers of which one specifically recognised mutated sequences introduced by the homology repair template and the other annealed to a genomic sequence adjacent to the area targeted for repair. Cell lines that appeared positive for the desired mutations from the genotyping PCR were then sequenced across the entire target site and its borders to confirm the presence of the correct mutation only.

2.3 Neural differentiation

Neural differentiation was carried out following a protocol adapted from that developed by Bibel et al. (Bibel et al., 2007). Embryoid body (EB) formation was induced by removal of LIF, on day 0 ESCs were plated onto non-adherent bacterial plates (Greiner) in 10 ml EB medium: DMEM containing 4.5 g/l glucose (Gibco) with 10% FBS (Life Technologies, South American, A3160802), 2 mM l-glutamine, 1 mM pyruvate, 1x MEM nonessential amino acids, 50 units/ml penicillin and 50 µg/ml streptomycin (all Gibco), and 0.2 mM -mercaptoethanol (Sigma), at a density of $4x10^6$ cells per 10 cm plate. Embryoid bodies began to form after one day. On day 2 the media was changed and increased to 13 ml by transferring cells to a 15 ml falcon tube, allowing EBs to settle at the bottom, and carefully aspirating the supernatant, and then adding fresh media before gently resuspending EBs and transferring to a fresh plate. On day 4 media was changed in the same manner to 15 ml of EB medium with 10 µM all trans retinoic acid (Sigma). On day 6 media was changed again maintaining the same volume of EB medium + retinoic acid. On day 8 the EBs were washed 3x in 20 ml PBS, trypsinised for approximately 3 min at 37°C with agitation and then quenched in 10 ml EB medium. Cells were then centrifuged for 5 min at 300 x g, resuspended in 10 ml EB medium and passed through a 40 µm cell sieve to remove clumped cells and undigested EBs. Cells were then counted, centrifuged again at 300 x g for 5 min and plated onto poly-D-lysine (Sigma) and laminin (Sigma) coated 6 cm dishes in 3 ml Advanced DMEM/F12 (Gibco) containing 1x N2 supplement (Gibco) at a density of $1x10^6$ cells/ml. On day 9 half the medium was replaced with Neurobasal medium (Gibco) containing 1x B27 supplement (Gibco), this was repeated on day 12 and day 15.

2.4 Gene expression analysis methods

2.4.1 RNA purification

RNA was extracted from cells by first resuspending in 1 ml TriPure RNA isolation reagent (Roche) and leaving samples at room temperature for 5 min. Samples were then centrifuged at 12,000 x g for 10 min at 4°C and the supernatant transferred to a fresh tube. 200 µl of chloroform was added to samples and mixed

by shaking before centrifuging samples again at 12,000 x g for 15 min at 4°C. The aqueous phase was transferred to a fresh tube along with an equal volume of isopropanol and 1 µl of glycoblue (Ambion). Samples were then centrifuged at 20,000 x g for 20 min at 4°C and the supernatant discarded. The pellet was then washed in 1 ml 70% ethanol then 1 ml 80% ethanol, centrifuging samples for 5 min at 20,000 x g at 4°C between washes. Supernatant was discarded and pellets allowed to air dry before resuspending in BTE (10 mM bis-tris pH6.7, 1 mM EDTA). Samples were treated with 2 units of Turbo DNAse (Ambion) for 30 min at 37°C before adding 1 ml of TriPure and 200 µl of chloroform. Protocol was then repeated up to the air drying of the RNA pellets which were resuspended in 50 µl BTE and concentration measured on Nanodrop. Quality of the RNA samples was further assessed on an Agilent bioanalyser using an RNA 6000 Nano chip. Samples were stored at -80°C.

2.4.2 cDNA synthesis and qPCR

cDNA samples were prepared from starting material of 1 µg RNA to which was added a final concentration of 2.5 µM oligodT and 0.5 mM dNTP mix. Samples were incubated at 65°C for 5 mins to anneal primer to RNA then returned to ice. To the annealed RNA/primer solution was added 1 µl, or 200 units of SuperScript IV reverse transcriptase and 4 µl 5X Superscript IV buffer with RNAase inhibitor and 100 mM DTT. Samples were transferred to thermocycler for 10 minutes each at 42°C, 50°C, 55°C and 80°C in succession. The resulting cDNA samples were then diluted at least 1/30 in water before being used in qRT-qPCR experiments and stored at -20°C. qRT-PCR experiments were carried out on a Lightycler 480 (Roche) using a 10 µl reaction volume with 2 µl of cDNA, 5 µM primers and 1x SyGreen Blue Mix (PCRBiosystems). Results were analysed using the 2-ΔΔCT method with *Gapdh* as the reference gene. Statistical analysis of the data generated was performed using a one-way ANOVA test followed by a Tukey Honest Significant Difference test.

2.4.3 RNA library preparation and quality control

RNA libraries were prepared from 1 µg of starting material as measured by Qubit (Invitrogen), using the NEBNext rRNA depletion kit and NEB Ultra II Directional RNA library prep kit for Illumina. All samples were first run on

Agilent bioanalyser RNA Nano 6000 chips to verify that all starting material had an RNA integrity number (RIN) of 8 or higher. Samples were then processed following directions provided with the kits. Briefly, rRNA depletion was carried out first by hybridising ssDNA probes to rRNA then treating with RNAse H which specifically degrades DNA/RNA hybrids. Samples were then treated with DNAse I to remove residual ssDNA probes before RNA was purified using Agencourt RNAClean XP beads (Beckman Coulter). Fragmentation of RNA was then performed for 15 mins at 94°C before first and second strand synthesis of cDNA were carried out. The double stranded cDNA was then purified using SPRIselect beads. End prep of the cDNA was then performed before ligation of adaptors and excision of the uracil by treatment with USER enzyme, breaking the hairpin structure formed by the adaptors. SPRIselect beads were then used to clean up the ligation reaction before proceeding to the PCR enrichment of the adaptor ligated cDNA using the universal PCR primer and a unique index primer for each separate library (NEBNEXT multiplex oligos for Illumina were used). Amplification was performed for 7 PCR cycles and samples purified with SPRIselect beads (Beckman Coulter). Quality and size distribution of the libraries was verified on Agilent bioanalyser DNA HS chips. Quantification of libraries was performed by qubit measurement and qPCR using the NEBNext Quant Kit for Illumina. Appropriate volumes of each samples were pooled to achieve a similar concentration of each library in the final pool, then submitted to the Edinburgh genomics facility for Next-Generation sequencing on an Illumina NovaSeq using an S2 flowcell and a 50 pair end read setup.

2.4.4 RNA-seq analysis

Once the raw data was received from the Edinburgh genomics sequencing facility, analysis was performed under my direction by Shaun Webb using the following methods.

Read processing and alignment

Preprocessing of reads was performed using Trimmomatic version 0.36 to trim low quality bases and adapter sequences from the ends of 50 bp, paired-end reads. Reads were then aligned to the mouse mm10 reference genome with STAR version v2.5.3.a. Properly paired primary alignments were selected for

downstream analysis and further filtering was performed to remove alignments to mm10 blacklisted regions. Unique alignments were obtained by selecting those with a mapping quality ≥20 using Samtools v1.9. Read coverage profiles for both strands were generated using bedTools v2.27 genomeCoverageBed to produce visualisation tracks.

Gene level quantification and differential expression analysis

Pseudo alignment of reads to transcripts was performed using Salmon v0.13.1 and transcript annotation, including non-coding RNA, from the Ensembl GRCm38.78 release. Read counts at the gene level were aggregated using the R package tximport and differential expression analysis was performed between multiple samples using DESeq2.

Gene ontology and transcription factor motif analysis

The GProfiler tool (biit.cs.ut.ee/gprofiler) was used for gene ontology analysis. Genes found to be significantly differentially expressed were grouped by cell line, timepoint and direction of change of expression (up or down). These groups were then analysed by GProfiler for the enrichment of terms associated with gene IDs using the Gene Ontology database which is subdivided into molecular function, cellular component and biological process. This tool was also used to interrogate the Kegg database which gives information on the biological pathways enriched amongst the genes of each group, as well as the TRANSFAC database which provides the regulatory motifs enriched in these genes. The transcription factor motifs were restricted to a +/- 1kb region around the TSS of genes.

2.5 DNA gel electrophoresis

All DNA samples were run at 100V on 1-2% agarose gels in TBE (100 mM Tris-HCl, 100 mM boric acid, 2 mM EDTA pH8) stained with SYBR safe (Invitrogen 1:50,000 dilution used). Gels were then imaged using the Chemidoc touch imaging system (Bio-Rad).

2.6 Western Blot

Protein samples were prepared by boiling whole cell pellets at 95°C for 5 min in SDS sample buffer (190 mM Tris-HCl, 30% glycerol, 6% SDS, 150 mM DTT, 0.3% bromophenol blue) using a ratio of 50 µl buffer per 1 million cells. Samples were then stored at -20°C. Proteins were separated by SDS-PAGE. Samples were loaded onto a1.5 mm mini acryalamide gel of appropriate percentage (15% for histone modifications, 8.5% for PRC2 subunits) and run at a voltage of 120 V for the stacking and 200 V for the separation gel. Cassettes used to cast gels and electrophoresis tanks from Novex by LifeTech. Proteins were then transferred to nitro-cellulose membrane using the semi-dry Trans-Blot turbo transfer system by Bio-Rad. Membranes were then blocked for 1 hour in a solution of 5% dry skimmed milk (Premier Foods) in TBS-t (20 mM Tris-HCL, 137 mM NaCl, 0.1% (v/v) Tween). Membranes were then washed 3x 5 min in TBS-t before adding primary antibodies, diluted appropriately in 5% BSA in TBS-t. This was incubated overnight with shaking at 4°C. Primary antibodies were then removed and membranes washed again 3x 5 min in TBS-t before adding corresponding HRP-conjugated secondary antibodies diluted 1:5000 in 5% BSA in TBS-t. This was incubated for 1 h with shaking at room temperature before removing secondary antibodies, washing 3x 5 min in TBS-t and developing membranes using the Clarity Western ECL substrate (Bio-Rad) and the Chemidoc touch imaging system (Bio-Rad).

2.7 Immunofluorescence staining of cells

2.7.1 Plating and fixation of cells

In preparation for plating 16 mm coverslips were soaked overnight in ethanol before being air dried then washed twice in PBS before coating. Cells were plated onto coverslips, coated with 0.1% gelatin (for ESCs) or with PDL and laminin (for NPCs) in 12 well plates. Cells were fixed either one day after being plated (for ESCs) or at different days of differentiation (for NPCs) by washing once in PBS and adding 4% paraformaldehyde for 20 mins at room temperature. Coverslips were then washed 3x 10 mins in PBS then stored in PBS at 4°C and used within 2 months of fixing.

2.7.2 Staining

Coverslips were washed once with PBS and blocked with 10% donkey serum in PBS with 0.1% Triton X-100 and incubated with primary antibodies overnight at 4°C. The following day, coverslips were washed 3x with PBS for 10 min, then incubated with Alexa fluor secondary antibodies (Invitrogen) at a dilution of 1:1000 in 1% donkey serum in 0.1% Triton X-100 for 1 h at room temperature in the dark. The coverslips were then washed 3x with PBS for 10 min. Nuclei were counterstained with 50 nM DAPI for 5 min, and washed 2x 10 min with PBS before being mounted on slides using vectashield.

2.7.3 Acquisition and processing of images

Imaging was carried out using a Zeiss Axio imager with 40x objective using standard filter sets. The microscope was equipped with a Hamamatsu Flash sCMOS camera. Micro-manager software (version 1.4) was used to capture images (Edelstein et al., 2014).

2.8 Chromatin immuno-precipitation

2.8.1 Chromatin preparation and quality control

Media was removed by aspiration from cells cultured on 15 cm plates and 15 ml fixation buffer (DMEM with 1% formaldehyde, 10 mM HEPES pH 7.6, 15 mM NaCl, 0.15 mM EDTA, 0.075 mM EGTA) was added. Crosslinking was carried out for 10 min at room temperature on a rocker. The formaldehyde was then quenched with 750 µl of 2.5 M glycine. The fixation buffer was then removed and cells were rinsed in 15 ml of cold PBS, resuspended in 2.5 ml cold PBS and pelleted at 2500 x g for 5 min at 4°C. The supernatant was then removed and the cell pellet weighed. The cells were resuspended in lysis buffer 1 (50 mM HEPES pH 7.6, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% Igepal, 0.25% Triton-X) to a concentration of 75 mg cells/ml. This was then placed on an end-over-end rotator for 10 min at 4°C and pelleted at 3000 x g for 5 min at 4°C. The supernatant was removed and the pellet resuspended in the same volume of lysis buffer 2 (10 mM Tris-HCl pH 8.0,200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA)

and again rotated at 4°C for 10 min before being pelleting at 3000 x g for 5 min at 4°C. The supernatant was removed and the pellet resuspended in the same volume of lysis buffer 3 (10 mM Tris-HCl, 1 mM EDTA, 0.5 mM EGTA, 0.5% N-lauryl sarcosine). Chromatin was sheared in a Bioruptor (Diagenode) for 17 cycles (30 s on, 30 s off) at the highest setting. The samples were centrifuged at 20,000 x g for 10 min at 4°C then the supernatant was aliquoted and frozen at -80°C. Size of chromatin fragments was verified on a 1% agarose gel and on a high sensitivity DNA chip in a 2100 Agilent bioanalyser after decrosslinking overnight at 65°C with shaking in elution buffer (100 mM $NaHCO_3$, 1% SDS, 200 mM NaCl, 0.4 mg/ml each proteinase K and RNAse A).

2.8.2 Immuno-precipitations

Chromatin was incubated overnight with antibodies in IP buffer (2% Triton X100, 200 mM NaCl, 1 mM PMSF, 1 mM EDTA, 10 mM Tris-HCl pH 8.0) rotating at 4°C, at this stage 10% input sample was taken from each IP. 50 µl per sample magnetic protein A or G DYNA bead suspension (Invitrogen) were washed twice using a magnetic rack (Invitrogen), then blocked overnight with PBS + 0.5% (w/v) BSA overnight at 4°C with rotation. On day 2 the beads were washed 5 times in TE buffer (10 mM Tris pH8, 1 mM EDTA) then aliquoted in protein low-bind tubes. IPs were spun at 20,000 x g for 10 min at 4°C to remove any precipitate that may have formed overnight. The supernatants were then transferred to the aliquoted DYNA beads and incubated at 4°C for 2-3 h rotating. Beads were then washed 5 times in cold RIPA buffer (10 mM HEPES pH 7.6, 1 mM EDTA, 0.5 M LiCl, 1% Igepal NP-40, 0.1% N-Lauryl Sarcosine, 0.2 mM PMSF) with 1-2 min rotating incubation for each wash. Beads were then washed once in TEN buffer (TE with 50 mM NaCl) without incubation. 200 µl elution buffer (100 mM $NaHCO_3$, 1% SDS, 200 mM NaCl, 0.4 mg/ml each proteinase K and RNAse A) was then added to beads and to samples corresponding to 10% of the input chromatin used per IP before incubating at room temperature for 15 min then at 65°C overnight with shaking to reverse crosslinking. The next day samples were purified using NEB Monarch columns and eluted using 50 µl kit elution buffer then stored at -20°C.

When using *Drosophila* S2 chromatin spike-ins in the preparation of samples for ChIP-seq, 100 ng of S2 chromatin (generated in house using the protocol outlined above) was added to each sample. During the immuno-precipitation

stage, an additional antibody for H2Av (Active Motif), a *Drosophila* specific histone variant, was included.

2.8.3 ChIP-qPCR

Eluted DNA was used in qPCR reactions on a Lightcycler 480 (Roche) in the same manner as described above. Enrichments were calculated as percentage of Input. Statistical analysis of the data generated was performed using a one-way ANOVA test followed by a Tukey Honest Significant Difference test.

2.8.4 ChIP library preparation and quality control

All ChIP samples were measured using qubit then 2-3 ng were used as starting material for library prep. Libraries were generated using the NEBNext Ultra II DNA library prep kit for Illumina following instructions provided with the kit. Briefly, end prep of DNA fragments was performed before ligation of adaptors, treatment with USER enzyme and 8 cycle PCR enrichment of libraries. Again, NEBNEXT multiplex oligos were used for index primers. Quality and fragment length of resulting libraries was assessed by bioanalyser using Agilent DNA HS chips. Concentration of each library was then determined by qPCR and by qubit measurements. Libraries were then pooled and submitted to Edinburgh genomics for Next-Generation sequencing on a HiSeq S1 flow cell using 50 paired end reads.

2.8.5 ChIP-seq analysis

Once the raw data was received from the Edinburgh genomics facility the analysis was performed under my direction by Shaun Webb using the following methods.

Read processing and alignment

Preprocessing of reads was performed using Trimmomatic version 0.36 to trim low quality bases and adapter sequences from the ends of 50 bp, paired-end reads. Reads were then aligned to a reference sequence containing the mouse mm10 and *Drosophila* dm6 assemblies, with bwa mem version 0.7.5.a using the -M option. Properly paired primary alignments were selected for downstream

analysis and further filtering was performed to remove alignments to mm10 blacklisted regions. Unique alignments were obtained by selecting those with a mapping quality ≥20 using Samtools v1.9, and duplicate reads were removed with Picard MarkDuplicates v2.20.3. Alignments were next split into mm10 and dm6 files to perform quantification and to calculate normalisation scaling factors. Spike in normalisation was performed following the protocol described at activemotif.com/documents/1977.pdf. This protocol is briefly described as follows: Uniquely aligning *Drosophila* sequencing tags were counted and the sample with the fewest tags was identified. By comparing this sample with each other sample a normalisation factor was generated for each sample. Final counts were acquired by multiplying each sample by its normalisation factor.

Read coverage and visualisation

Read coverage profiles as bigWig files were generated using Deeptools v3.13 bamCoverage and bamCompare.

2.9 Fluorescence *in situ* hybridisation

2.9.1 Probe preparation

Fosmid probes were selected from the WIBR-1: Mouse Fosmid Library (Whitehead Institute/MIT). Probes were prepared from glycerol stocks stored at -80°C. Probes are listed in Fig.2.1. Probes were prepared by small scale DNA plasmid preparation (Omgea) and stored at -20°C. Probes were directly labelled with fluorescent nucleotides by nick translation, using the following reaction mix (20 µl): 2 µl Nick translation salts, 2.5 µl 0.5 mM dATP, 2.5 µl 0.5 mM dCTP, 2.5 µl 0.5 mM dGTP, either 2.5 µl 1 mM ChromaTide Alexa 594-5-dUTP (Invitrogen) or 2.5 µl 1 mM 5(6)-Carboxyrhodamine Green dUTP (Enzo), 6 µl fosmid miniprep DNA (estimated 0.5-1ug), 1 µl DNase I (1:10 dilution) and 1 µl DNA polymerase I (Invitrogen). This reaction was incubated for 90 mins at 16°C. The reaction was stopped by adding 3 µl 0.5M EDTA (pH 8) and 2 µl 20% SDS. 65 µl of TE buffer was then added to the sample, before successfully labelled DNA was purified using a Quick spin column (Pharmacia) per manufacturers instructions. Labelled fosmid DNA was then stored at -20°C.

Locus	Whitehead (Sanger) Name	Coordinates (mm9)	Size (bp)
LHX5 - PE	WI1-2840O15	Chr5 120769676 - 120808972	39297
LHX5- Promoter	WI1-0129N23	Chr5 120857793 - 120898771	40979
LHX5 - Control	WI1-2079G02	Chr5 120945914 - 120986094	40181

Figure 2.1 FISH probe location and size

2.9.2 Preparation of cells

Superfrost plus slides (Thermo Scientific) were prepared by soaking overnight in 70% ethanol then air drying and washing in PBS before plating cells at approximately 1-2x10^6 cells per slide. The following day, slides were washed 3x in PBS before being fixed in 4% paraformaldehyde for 10 min. Slides were then washed 3x in PBS and treated with 0.5% Triton X-100 for 10 min to permeabilise cells before being washed again 3x in PBS. Slides were finally allowed to air dry before storing at -80°C.

2.9.3 Probe hybridisation

Slides were briefly washed in 2x saline-sodium citrate (SSC) buffer, then incubated for 1 hour in 2x SSC with 100 µg/ml RNaseA at 37°C. Slides were then washed in 2x SSC before cells were dehydrated using a series of ethanol washes (2 min each 70%, 90% then 100% ethanol). Slides were air dried and then heated to 70°C in an oven for 5 min before denaturation in 2x SSC with 70% formamide (Honeywell), pH 7.5 at 80°C for 30 min. A second dehydration step was then performed (2 min each 70% ice cold, 90% room temp, 100% room temp ethanol). Slides were air dried then incubated at 37°C while probes were prepared.

For each slide 10 µl of directly-labelled probe was used as well as 8 µg of mouse Cot1 DNA (Invitrogen) and 5 µg sonicated salmon sperm DNA, used to prevent non-specific binding during hybridisation steps. 2x volumes of 100% ethanol were used to precipitate probes which were then vacuum dried then 15 µl hybridisation solution (50% deionised formamide, 10% dextran sulphate, 1% Tween-20 in 2x

SSC) was added to the precipitated probes, and allowed to dissolve for 1 h at 30°C. Denaturation of the probes was then carried out at 80° for 5 min followed by pre-annealing at 37°C for 15 min. Probes were then added to a 22 mm × 22 mm coverslip at 37°C before the slides that had been pre-warmed at 37°C were placed face down on coverslips. A rubber solution was used to seal the coverslips and slides which were then left to hybridise overnight at 37°C. Following hybridisation, the rubber seal was removed and slides were washed 4x in 2x SSC for 3 min at 45°C and then 4x in 0.1x SSC for 3 min at 60°C. Slides were then washed once in 4x SSC with 0.1% Tween-20 before adding 50 µg/ml DAPI in 4x SSC with 0.1% Tween-20 for 3 min. A 22 mm × 40 mm coverslip is then added with Vectasheild mounting medium (Vector labs) and sealed with nail varnish.

2.9.4 Acquisition and analysis of images

FISH images were acquired using a Zeiss Axio imager with 100x objective using standard filter sets. The microscope was equipped with a Hamamatsu Flash sCMOS camera and a Marzhauser 8 slide motorised stage. Images were taken with a 0.2 µm z-step. Micro-manager software (version 1.4) was used to capture images (Edelstein et al., 2014).

The 3D captured images were then deconvolved using Autoquant X3 software with blind 3D deconvolution settings. Imaris 8.0 was used for image analysis, where the Spots function was used to mark hybridised probe signals, and the XYZ coordinates of the centre of each probe was recorded. The 3D distances between each red probe and its closest green pair were calculated using an automated javascript written by Dr David Kelly. The results were then manually assessed to select only the true probe pairs, and to discard any aberrant signals and ensure data was reliable. Measurements were made for at least 80 cells per genotype across at least 2 slides. Statistical significance of the differences between measurements were determined by performing a Kruskal-Wallis test followed by a Mann-Whitney test.

Chapter 3

Generation and characterisation of EZH2 knockout and knockin cell lines

3.1 Introduction and aims

In order to address the main question of my project I first aimed to establish a system in which I could compare the absence of both PRC2 and H3K27me3 from chromatin to the loss of H3K27me3 alone while maintaining an intact PRC2 complex. By using mESCs I was able to examine the effects of a catalytically inactive PRC2 on the chromatin state and gene regulation of its target genes not only in a steady state but also in the dynamic transition through differentiation. Bivalency was also first identified in mESCs in 2006 and this cell type has been used extensively for the study of Polycomb group proteins. In this way, any findings from this project will be directly comparable to many published studies, facilitating the positioning of my work in the context of the broader field of Polycomb research. In light of this we chose to generate mESC lines with either a catalytically inactive PRC2 or a PRC2 KO. The gene editing method we chose for introducing these mutations was CRISPR Cas9 genome editing, due to its simplicity in design, ease of implementation, and high specificity. In order to obtain all PRC2 mutants in the same genetic background and to minimise differences between the lines due to varied time in culture and gene editing methods, I opted to generate all cell lines myself rather than using pre-existing

lines with different backgrounds, for which it would be difficult to control.

Many previously generated EZH2 KO cell lines have used C-terminal truncations which, while indeed leading to loss of the catalytic function of the protein, does not necessarily lead to complete removal of the EZH2 protein, dissociation of the complex and thus loss of its binding to chromatin, nor would it ensure the loss of any potential activity of EZH2 that is distinct from the enzymatically active C terminus (Shen et al., 2008; Højfeldt et al., 2018). In this project I aimed to block the production of the entire protein in order to not only lose the methylation activity of PRC2 but also to disrupt potential other functions located within the N terminus as well as the assembly of the EZH2-PRC2 complex and consequently its binding to chromatin. Moreover, I also chose to target EZH2 itself rather than one of the other core subunits to minimise the differences between the KO mutant and the catalytic mutant to ensure that any differences that I may observe between the two lines are not a consequence of secondary effects due to the loss of the SUZ12 or EED subunits. This strategy also allowed me to study EZH1 in the absence of EZH2 with the aim of gaining further insight into their distinct functions.

To generate a catalytically inactive form of EZH2 I aimed to introduce as few mutations as possible to selectively target the active site of the SET domain without unintentionally disrupting other functions of the protein. It has been shown that within the SET domain of all protein lysine methyltransferases there are 2 conserved tyrosines, the mutation of one of which has been shown to completely abrogate methyltransferase activity (Kwon, 2003). In EZH2 this corresponds to tyrosine 726, one of several aromatic residues located in the methyl-lysine binding pocket of the protein (Kipp et al., 2013). We chose to mutate this highly conserved residue in order to achieve robust inactivation of the catalytic activity of EZH2.

In this chapter I will describe in detail how we designed and carried out the gene editing of the mESCs as well as the validation methods used to confirm that the cell lines were indeed mutated at the desired sites and produced the expected effect on EZH2 and H3K27me3 levels.

The work in this chapter was performed with the help of several members of the Voigt lab under my direction. Genotyping and passaging of the EZH2 Y726A and EZH2 KO cell lines was done with the help of Elana Bryan and the cloning of the pX458 plasmid targeting EZH1 was done by Viktória Major, while Katy

McLaughlin and Giulia Bartolomucci helped with the genotyping of these cell lines.

3.2 EZH2 single mutants

I initially chose to target the EZH2 subunit of PRC2 as its expression is associated with highly proliferating cells and thus is much more highly expressed in ESCs than EZH1 (Laible et al., 1997; Bracken, 2003). It has also been shown that EZH2 is more enzymatically active than its paralogue and responsible for the vast majority of H3K27me3 in ESCs (Margueron et al., 2008). I therefore targeted the EZH2 subunit in order to achieve the strongest reduction in H3K27me3 levels in ESCs, expecting the remaining H3K27me3 levels due to EZH1 activity to be negligible.

3.2.1 Design of EZH2 CRISPR targeting strategies

Each genotype was generated using a similar CRISPR targeting strategy. gRNAs were generated by the CRISPR design tool created by the Zhang lab at MIT (crispr.mit.edu now defunct) and a suitable gRNA was chosen based on proximity to the bases to be altered in the genomic sequence. Once optimal gRNAs for the target locus were identified, a 200-bp ultramer to be used for homologous repair of the target site was designed. This single-strand donor oligonucleotide (ssODN) DNA fragment included the desired mutations as well as additional mutations within the PAM and/or gRNA hybridisation site to prevent cleavage of the repair template and the repaired target sites by the Cas9 endonuclease. For the KO of EZH2 we aimed to introduce premature in-frame stop codons into an early exon that would result in a truncated and presumably unstable protein fragment. Alternatively, the stop codons could induce non-sense mediated decay of the transcripts generated from the gene and in turn lead to a loss of protein expression. We initially attempted this at exon 2 and exon 3 of the EZH2 gene. Although in both cases clones with correctly mutated DNA sequences were obtained, the mutations did not result in loss of protein expression. We found that with both targeting strategies EZH2 was still detectable by Western blot but appeared to be migrating at a slightly smaller size than expected, indicating that the mutations introduced by CRISPR may have caused these exons to be skipped

resulting in a N-terminally truncated EZH2 protein. Finally, our third attempt targeting exon 7 was successful in producing a KO of EZH2. The ultramer designed for this targeting approach contained the mutations shown in Fig. 3.1. To generate a catalytically inactive form of EZH2 we targeted the SET domain located in part in exon 19. Here we designed the ultramer to replace tyrosine 726 with an alanine (Y726A), while also introducing silent mutations in the gRNA hybridisation site to prevent spurious cleavage, as shown in Fig. 3.2.

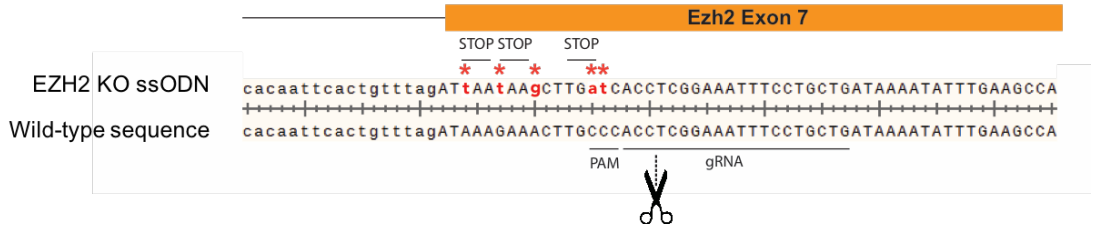


Figure 3.1 Design of EZH2 KO CRISPR targeting strategy. Alignment of E14 Wild-type sequence of exon 7 of *Ezh2* gene with the ssODN designed to induce EZH2 KO mutations (uppercase is exonic sequence, lowercase is intronic). gRNA hybridisation site and PAM sequence are labelled, Cas9 cut site is indicated with scissors. Point mutations in ssODN are indicated by red asterisks (lowercase in red), in frame stop codons are labelled.

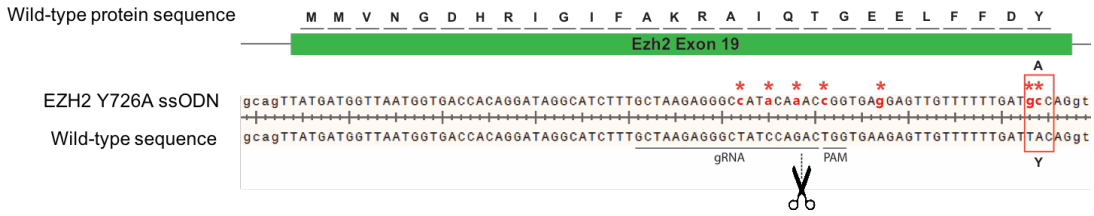


Figure 3.2 Design of EZH2 Y726A CRISPR targeting strategy. Alignment of E14 Wild-type sequence of exon 19 of *Ezh2* gene with the ssODN designed to induce EZH2 Y726A mutations (uppercase is exonic sequence, lowercase is intronic). gRNA hybridisation site and PAM sequence are labelled, Cas9 cut site is indicated with scissors. Point mutations in ssODN are indicated by red asterisks (lowercase in red), critical Y726A mutation is highlighted. Wild-type protein sequence is displayed above corresponding codons.

3.2.2 Screening and validation of successful targeting

Low passage E14 mESCs were first transfected with the repair ssODN and the pX458 plasmid from which the gRNA and Cas9 protein are expressed. Cas9 is expressed as a 2A peptide-cleavable fusion with GFP, which serves as a reporter

of cells that have been successfully transfected and are expressing Cas9. 48 hours after transfection, cells were sorted by FACS and single GFP positive cells were plated sparsely and grown before being harvested as single colonies onto 96-well plates for expansion and genotyping. The strategy used for genotyping (shown in Fig. 3.3) was a PCR screen using a mutation-specific primer along with a primer situated outside of the mutation site and the sequence covered by the ssODN. In untargeted ESCs these primer pairs did not allow for correct amplification and a PCR product for the predicted size could only be observed in the presence of the correct mutated sequence at the target site. From this mutation-specific PCR potentially positive clones were chosen for a secondary PCR screen using primers situated in flanking regions of the mutation site allowing amplification of a DNA fragment spanning the entire ssODN-targeted region. This PCR product was then sequenced using both the forward and reverse genotyping primers to ensure coverage across the target site. An example of this screening process is shown in Fig. 3.4, with an example of a successfully homozygously targeted EZH2 Y726A clone shown in Fig. 3.5. Using this strategy, we were able to identify several clones of each genotype with the correct mutations both in a heterozygous and homozygous state.

Figure 3.3 Schematic representation of genotyping PCR strategy. In blue, forward and reverse genotyping primers annealing to flanking regions are labelled FG and RG. In red, the mutation specific reverse primer is labelled RM. Induced mutations in targeted sequence are represented by red asterisks.

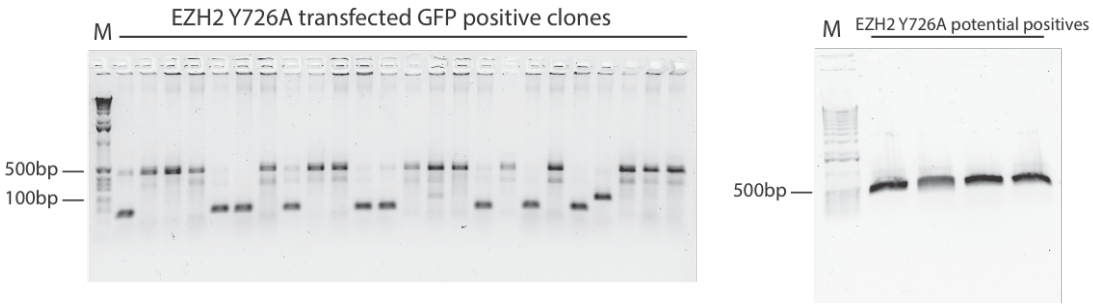


Figure 3.4 Example of CRISPR genotyping PCR. Mutation specific genotyping PCR of EZH2 Y726A clones generates a PCR product of around 70 bp when the correct mutations are present at target site. Outside genotyping primers amplify a fragment of around 500 bp to be extracted and sequenced.

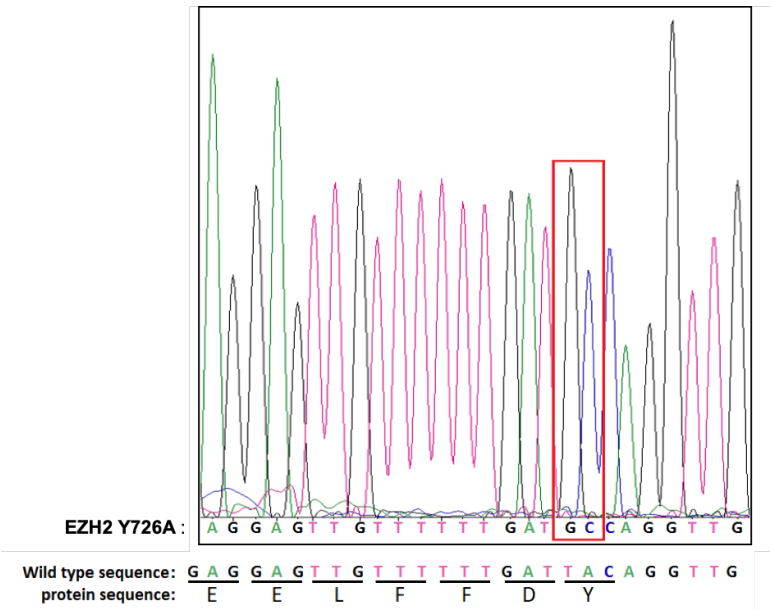


Figure 3.5 Sequence of a homozygous EZH2 Y726A clone. Example of a sequencing chromatogram for EZH2 Y726A showing the correct mutation of the nucleotides within the codon for the catalytic tyrosine of the EZH2 SET domain.

3.2.3 Phenotypic validation of successfully targeted cell lines

Once all potential positive clones were validated by sequencing each cell line was expanded and subjected to cryo-preservation before performing further screening at the protein level. First, I performed Western blot on whole cell extracts to establish the presence or absence of EZH2 in the mutant cell lines (see Fig. 3.6). In the EZH2 KO cell lines I confirmed that the protein was no longer detectable in the homozygous clones, whereas in the heterozygous clones I observed a reduction of EZH2 as compared to control E14 mESCs. Next I analysed whether the Y726A mutation affected the expression of EZH2 in either heterozygous or homozygous cells. I found that the expression levels of EZH2 appeared to be slightly reduced in some clones but not others, as can be seen in Fig. 3.6. In the following chapters I chose to use only clones for which the level of EZH2 expression in the Y726A mutants was the closest to that of the E14 control mESCs.

Once I had confirmed the expected expression profiles of EZH2 I then wanted to verify that both EZH2 KO and EZH2 Y726A mutations resulted in comparable losses of H3K27me3. Both EZH2 Y726A and EZH2 KO clones were found to have diminished, although still detectable levels of H3K27me3 as compared to control E14 mESCs (see Fig. 3.7). An EED KO cell line (Chamberlain et al., 2008)

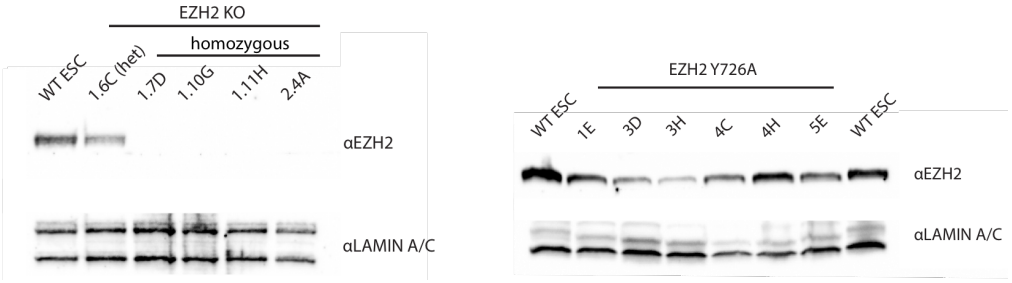


Figure 3.6 EZH2 expression in EZH2 KO and EZH2 Y726A cells. Western blot using an antibody against EZH2 on whole cell extracts of EZH2 KO and EZH2 Y726A clones confirmed by sequencing to have integrated the desired mutations. Lamin A/C was used as a loading control.

was used as a control in these blots as it has been shown that the knockout of this PRC2 subunit results in the complete loss of H3K27 methylation. Although the reduction of H3K27me3 observed in both EZH2 Y726A and EZH2 KO was similar, it was significantly higher than the levels detected in the EED KO.

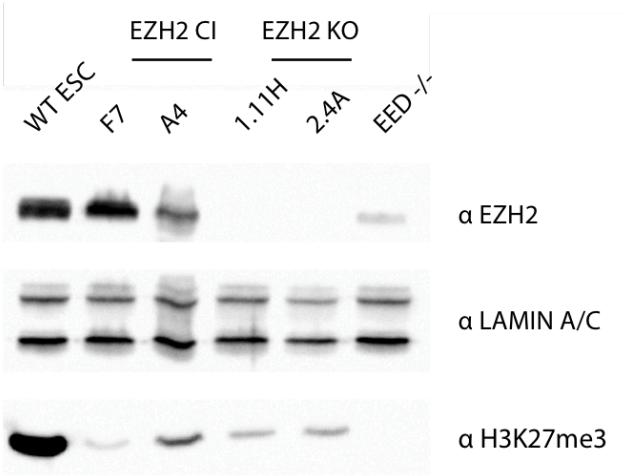


Figure 3.7 H3K27me3 levels in EZH2 KO and EZH2 Y726A cells. Western blot using an antibody against H3K27me3 on whole cell extracts of EZH2 Y726A (CI) and EZH2 KO clones. Lamin A/C was used as a loading control.

3.3 Generation of EZH1/2 double mutants

The reduction of H3K27me3 in these first cell lines was not as complete as I had predicted from the loss of EZH2 activity in light of the expected lower expression and activity of EZH1. I hypothesised that any role of EZH2-PRC2 that relies on H3K27me3 may still function even with the low level of residual tri-methylation

catalysed by EZH1. For this reason, I decided to also combine the knockout of EZH2 as well as the EZH2 Y726A mutation with the knockout of the less active, less abundant EZH1 subunit in these cells, in order to obtain a more complete loss of the H3K27me3 mark. In this way any potential catalysis-independent functions of EZH2-PRC2 can be assessed without confounding residual H3K27me3. In the following chapters work was carried out using both the EZH2 single mutant cell lines as well as the EZH1/2 double mutant cell lines.

3.3.1 Design of EZH1 CRISPR targeting strategy

The EZH1 KO targeting strategy was analogous to that used for the knockout of EZH2 within exon 7. As can be seen in Fig. 3.8, several point mutations were introduced in an early exon (exon 8) to produce in-frame stop codons. Additional point mutations were introduced in the gRNA hybridisation site to prevent repeated cleavage of the target.

Figure 3.8 Design of EZH1 KO CRISPR targeting strategy. Alignment of E14 Wild-type sequence of exon 8 of *Ezh1* gene with the ssODN designed to induce EZH1 KO mutations (uppercase is exonic sequence, lowercase is intronic). gRNA hybridisation site and PAM sequence are labelled, Cas9 cut site is indicated with scissors. Point mutations in ssODN are indicated by red asterisks (lowercase in red), in frame stop codons are labelled.

3.3.2 Screening and validation of successful targeting of EZH1

For the double mutants I initially attempted to introduce mutations targeting both *Ezh1* and *Ezh2* simultaneously in E14 mESCs. This would have allowed me to generate clones that were independent of the initial EZH2 single mutant clones that I had already created. However, this strategy resulted in high amounts of cell death and low levels of successful transfection (as measured by percentage of

GFP expressing cells detected by FACS), perhaps due to the increased amount of lipofectamine and DNA present in the transfection reaction. Ultimately, we were unable to obtain any correctly targeted EZH1/2 double mutants from this attempt. We were however able to obtain EZH1 single KO and additional, independent EZH2 KO and Y726A clones from this round of transfections. After this initial failed attempt to obtain the EZH1/2 double mutants we changed strategy and instead used two clones each of the EZH2 KO and EZH2 Y726A cell lines and transfected each of them with the EZH1 KO constructs. The screening was performed in the same way as the previous mutants by performing a genotyping PCR using a mutation-specific primer to identify the mutated clones before sequencing these clones across the target site to confirm the presence of the desired mutations in the *Ezh1* gene.

3.3.3 Phenotypic validation of successfully targeted cell lines

As previously described for the single mutants, once all clones were verified by sequencing the knockout of EZH1 was then further verified by Western blot (see Fig. 3.9). In both the EZH1/2 KO/Y726A and EZH1/2 dKO, EZH1 was at least strongly reduced in expression if not completely absent from each clone. Levels of H3K27me3 were also assessed by Western blot and all EZH1/2 double mutants were found to have levels of H3K27me3 that were no longer detectable.

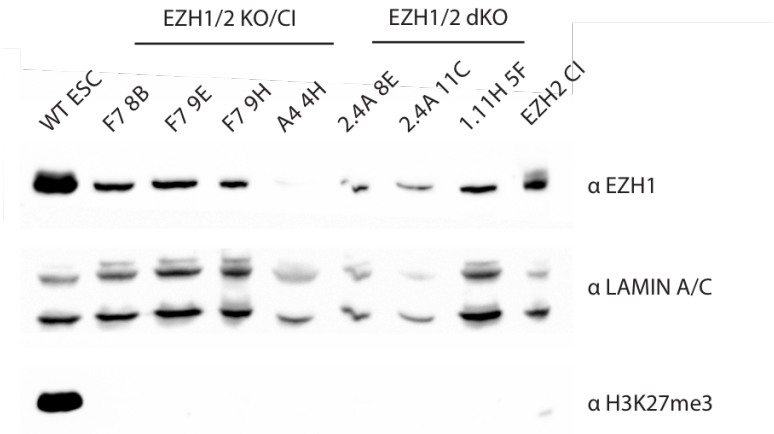


Figure 3.9 EZH2 and H3K27me3 levels in EZH1/2 KO/Y726A and EZH1/2 dKO cells. Western blot using antibodies against EZH2 and H3K27me3 on whole cell extracts of E14 control (WT ESC), EZH1/2 KO/Y726A (KO/CI) and EZH1/2 dKO clones confirmed by sequencing to have integrated the desired mutations. Lamin A/C was used as a loading control.

3.3.4 Effects of EZH1/2 mutations on overall PRC2 expression

Once the mutations generated in EZH1 and EZH2 were validated, I then verified by RT-qPCR whether these mutants had any effect on the mRNA expression of PRC2 complex members. As shown in Fig. 3.10, this experiment further confirmed the loss of expression of EZH2 in the EZH2 single KO and EZH1/2 dKO cell lines and of EZH1 in both of the EZH1/2 double mutant cell lines, whereas the EZH2 Y726A mutation did not affect the expression of EZH2 at the mRNA level compared to the E14 control. The mRNA expression of EZH1 was found to be largely unchanged following mutations of EZH2, however there did appear to be a slight upregulation of this gene in the EZH2 single Y726A cell lines. The EED subunit was slightly decreased in expression in the EZH2 single mutant cells, and even more so in the EZH1/2 dKO cells whereas it was largely unaffected in the EZH1/2 KO/Y726A cells. SUZ12 was expressed close to wild-type levels in the EZH2 single mutant cells but was expressed at lower levels in the EZH1/2 double mutant cell lines (see Fig. 3.10).

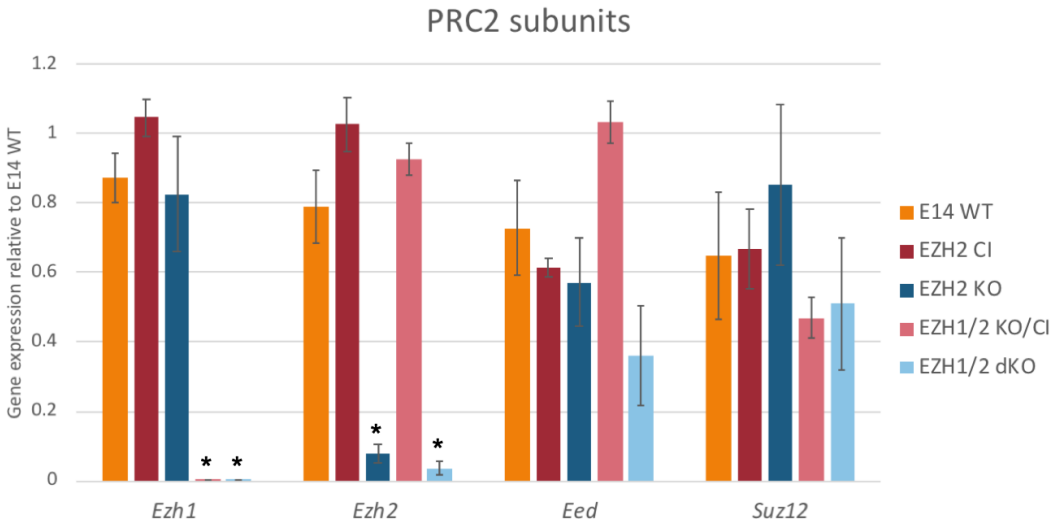


Figure 3.10 Expression of PRC2 subunits in EZH2 single and EZH1/2 double mutant mESCs. Bar chart showing expression of PRC2 subunits in EZH2 Y726A (CI), EZH2 KO, EZH1/2 KO/Y726A (KO/CI) and EZH1/2 dKO cell lines relative to E14 control. Data was generated by RT-qPCR and normalised to *Gapdh*. Error bars represent standard error of the mean across 3 independent clones, * denotes a significant difference to WT with a p value of less than 0.01.

3.4 Conclusions

I successfully generated all required PRC2 mutant cell lines using CRISPR Cas9 genome editing. I found that while the KO of EZH2 alone was successful in knocking out any detectable levels of the EZH2 protein and strongly reducing the mRNA expression of the gene, low but detectable levels of H3K27me3 persisted in these cell lines. The amount of H3K27me3 remaining also appeared to vary slightly between clones. In the EZH1/2 double mutants, although I found by Western blot that the EZH1 protein was not completely lost in most clones, by RT-qPCR I observed a complete loss of expression of the gene. These seemingly conflicting findings might be due to unspecific reaction of the EZH1 antibody with a protein of similar size to EZH1 itself. I further observed that the KO of EZH1 in addition to the mutations of EZH2 led to the complete loss of H3K27me3. I did not find any strong effects of the mutation of these two subunits of PRC2 on the expression of any of the other core subunits of the complex, although increased clonal variability in the expression of these subunits was observed in the EZH1/2 double mutants. Importantly the catalytically inactive form of EZH2 was found to still be able to associate as expected with the other subunits enabling the assembly of the full PRC2 complex.

In the following experiments in this project I used 3 independent clones of each genotype as biological replicates in order to minimise clonal effects due to the selection of individual single cell clones. In addition to heterogeneity between individual E14 mESCs, such clonal effects could arise from off-target activity during genome editing, which is known to be low with CRISPR Cas9 but not completely unavoidable. Each clone was confirmed to have the same sequence at the targeted sites but were each derived from distinct colonies.

Chapter 4

Characterisation of chromatin state in PRC2 mutant cell lines

4.1 Introduction and aims

After successfully generating all the mutant PRC2 cell lines I next aimed to investigate how the chromatin state of Polycomb-bound genes was affected. PRC2 is able to bind to chromatin through multiple interactions with the nucleosome involving each of the core subunits of the complex. EZH1 and EZH2 themselves are able to bind to the H3 tail by the interaction of lysine 27 with the binding pocket of the SET domain. PRC2 is also able to read its own mark by the recognition of H3K27me3 by the EED subunit. The remaining core subunits SUZ12 and Rbap46/48 have also been shown to interact with histone H3, an interaction which has been shown to be markedly weaker in the presence of H3K4me3 (Schmitges et al., 2011; Yuan et al., 2012). PRC2 is known to bind to promoters of its target genes, most of which in ES cells are involved in development, and has a distinct preference for genes whose promoters contain CpG islands devoid of DNA methylation (Boyer et al., 2006; Mikkelsen et al., 2007; Mendenhall et al., 2010). The pattern of PRC2 binding to these loci mostly follows a similar profile with relatively broad peaks centred around the TSS of the gene. The profile of H3K27me3 at these genes tends to closely mirror that of PRC2 binding. The allosteric activation of the PRC2 complex through the interaction with the EED subunit is thought to be responsible for these broad profiles across these genes. It has recently been suggested that PRC2 may initially

be recruited to its targets through "nucleation" sites from which the complex, and thus the histone PTM it places are then able to spread (Oksuz et al., 2018). It has also been proposed that this spreading could even happen in trans to secondary target genes through chromatin looping and not just in cis around the initial nucleation site within the same chromosome. In line with this discovery I was curious to find out whether the catalytically inactive EZH2 would have a binding pattern similar to that seen in this study. If the spreading of the PRC2 complex indeed occurs solely via the recognition of its own mark through the EED subunit, it is to be expected that the loss of catalytic activity would lead to a disruption of this spreading.

After successfully generating the mutant PRC2 cell lines, I next aimed to investigate how the chromatin state of Polycomb bound genes was affected. This chapter describes ChIP-qPCR experiments that were performed to characterise the chromatin state of the PRC2 mutant cell lines. With these experiments, I aimed to address the following questions: To verify that the catalytically inactive PRC2 complex was a suitable model to explore catalysis-independent actions of PRC2, I sought to determine whether it was still able to bind to chromatin in absence of H3K27me3 and in the presence and absence of EZH1. I also sought to determine the distribution of H3K27me3 that is maintained by EZH1. It has also been shown that while H3K4me3 and H3K27me3 coexist at bivalent genes, these two marks are mutually exclusive on the same histone tail and it has been shown *in vitro* that PRC2 is not able to modify histone tails already carrying H3K4me3 (Schmitges et al., 2011; Voigt et al., 2012). In light of this I also sought to determine whether the loss of H3K27me3 alone or along with PRC2 would have a considerable effect on the abundance of H3K4me3 at bivalent genes.

4.1.1 Changes in global histone PTMs

Before investigating the chromatin state at specific loci and genome-wide, I first assessed global levels of certain histone modifications to determine whether any of the mutations of PRC2 caused striking changes in modification levels. I also examined the expression of a selection of Polycomb group proteins to verify that the abundance of these components remained unchanged. Overall levels of histone modifications relevant to the PcG proteins and bivalent genes were determined by Western blot of whole cell extracts of the single EZH2 mutants (Fig. 4.1). In both EZH2 Y726A and EZH2 KO cell lines the H3K27me3 histone

mark was strongly reduced although not completely lost, as was seen in the EED KO cells, used here as a control. This was in line with the loss of EZH2 catalytic activity in both cell lines, either by mutation of the SET domain in the EZH2 Y726A lines, or by loss of the protein itself in the EZH2 KO lines. As described in the previous chapter, remaining H3K27me3 observed in these cell lines is likely to come from EZH1 which is known to be able to carry out the same catalytic functions as EZH2 but with much lower activity. Accordingly, the H3K27ac mark, which is mutually exclusive with H3K27me3, increased slightly in the EZH2 mutant cell lines to an extent that was inversely proportional to the loss of H3K27me3. The active histone mark H3K4me3 which co-occupies bivalent genes with H3K27me3, did not exhibit any strong global changes in abundance in the EZH2 single mutants although there was a subtle increase seen in the EZH2 Y726A cell lines. Surprisingly, there seemed to be an increase in the H3K27me1 levels in the EZH2 KO and one of the EZH2 Y726A cell lines. With regard to Polycomb group proteins, EZH2 was detected in EZH2 Y726A cells but not in EZH2 KO cells (Fig. 4.1, see also Fig. 3.7). Levels of the PRC2 subunit SUZ12 remained largely unchanged in all EZH2 mutant cell lines. Nor did I observe any striking changes in the PRC1 complex (Fig. 4.1). Neither RING1B nor CBX7, the PRC1 subunits responsible for placing H2AK119ub and binding to H3K27me3 respectively, exhibited altered abundance following the mutation of EZH1 or EZH2.

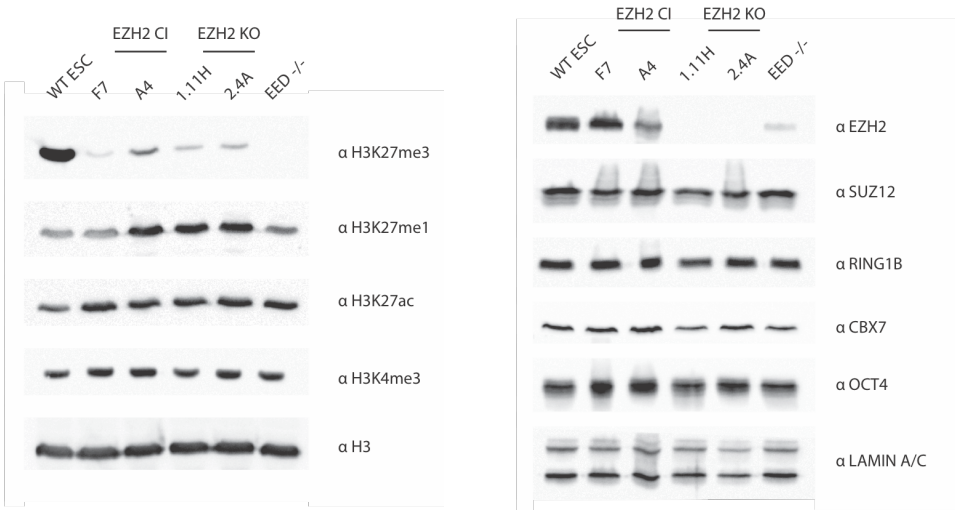


Figure 4.1 Global levels of histone PTMS and PcG proteins in EZH2 single mutant mESCs. Western blots analysis of histone PTMs and PcG proteins using whole cell extracts and indicated antibodies. Histone H3 and Lamin A/C were used as loading controls.

4.2 Analysis of chromatin state at bivalent genes by ChIP-qPCR

I next sought to determine the levels of binding of EZH2 to its target genes in the mutant cell lines and to asses the abundance of H3K27me3 and H3K4me3 at these genes by performing ChIP-qPCR. To determine whether the catalytically inactive form of EZH2 retained its chromatin binding activity I performed ChIP for EZH2 with each of the PRC2 mutant cell lines. I then probed for binding to regions that corresponded to the promoters of active genes (*Gapdh*, *Oct4*) and of known Polycomb bound genes (*Hoxc5*, *Hoxc11*, *HoxB13*, *Gata4*, *Bcor*, *Olig1*, *Pax3*) as well as to the gene body of a Polycomb bound gene (*Fabp7*). I also performed ChIP-qPCR in all PRC2 mutant cell lines for levels of H3K27me3 and H3K4me3 at the same sites to determine how their abundance is affected at bivalent genes in the absence of PRC2 activity.

4.2.1 Analysis of H3K27me3 levels at bivalent genes

I first assessed the levels of H3K27me3 at active and Polycomb target genes in E14 control and PRC2 mutant cell lines. In the E14 control H3K27me3 was found to be enriched to varying degrees at all Polycomb target genes as expected, while there was no significant enrichment at either of the active genes (Fig. 4.2). Interestingly, H3K27me3 was still detectable within the gene body of *Fabp7*, suggesting a degree of spreading of the mark from the promoter region of this gene. In both EZH2 single mutants there was a strong reduction of H3K27me3 at all Polycomb target genes. While it was depleted to a similar degree in both the EZH2 KO and EZH2 Y726A, I found that most Polycomb target genes retained low but detectable levels of H3K27me3 at their promoters. In the *Fabp7* gene body however, I found there to be no enrichment of H3K27me3 in either EZH2 single mutant. In the case of the EZH1/2 double mutants I found that H3K27me3 was completely depleted from all of the Polycomb target genes with no residual enrichments detected for any of the genes probed. The levels of H3K27me3 detected in this assay in both the EZH2 single and EZH1/2 double mutants reflected the global levels of this histone modification that I had previously detected by Western blot. This further confirmed that the EZH2 Y726A mutant was indeed devoid of catalytic activity and that the residual methylation of lysine 27 detected in the EZH2 single mutants was abolished by the knockout of EZH1.

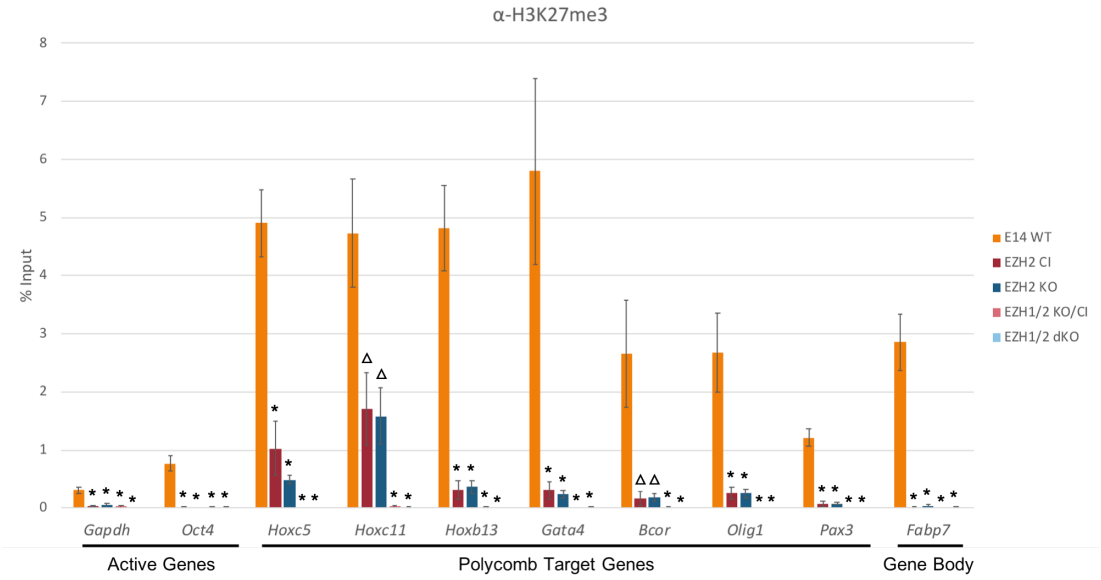


Figure 4.2 H3K27me3 enrichment in E14 control and PRC2 mutant mESCs. ChIP-qPCR for H3K27me3 in E14 control (WT), EZH2 Y726A (CI), EZH2 KO, EZH1/2 KO/Y726A (KO/CI) and EZH1/2 dKO cells. Enrichments were normalised to input, with error bars representing standard error of the mean across three biological replicates. * denotes a significant difference to WT with a p value of less than 0.01 and Δ a p value of less than 0.05.

4.2.2 Analysis of EZH2 binding to bivalent genes

I next sought to determine how the distribution of EZH2 was altered in the mutant cell lines. In the E14 control cells EZH2 was found to be enriched at Polycomb target genes, but was absent from the active genes (Fig. 4.3). The enrichment of EZH2 was found to broadly mirror the levels of H3K27me3 detected at all genes in the E14 control with the exception of the *Fabp7* gene body. Though this gene was found to be enriched for H3K27me3 in the E14 control, this was not the case for EZH2, indicating that while EZH2 may be more restricted to promoter regions, its mark can spread further into the gene body. In the EZH2 single and EZH1/2 double mutants the levels of EZH2 detected at active genes and the *Fabp7* gene body were comparable to that seen in the E14 control. At Polycomb target genes the catalytically inactive EZH2 Y726A was found to bind to varying levels, at some genes such as *Hoxc11* and *Bcor* this level was similar to that observed in the E14 control whereas at others such as *Gata4* and *Olig1* this enrichment was reduced by more than twofold. At the *Fabp7* gene body, as in the E14 control, EZH2 Y726A was not found to be enriched. In the EZH2

KO cell line EZH2 binding was found to be strongly reduced at all Polycomb target genes, although this depletion was not quite as complete as expected given that this protein was undetectable by Western blot. In the EZH1/2 KO/Y726A cells, though EZH2 binding was found to be reduced as compared to the E14 control, there was still an enrichment of this protein detected at all Polycomb target genes. Finally, in the EZH1/2 dKO cells, I found that EZH2 was virtually undetectable both at the active genes and the Polycomb target genes. Together the results of this assay showed that the catalytically inactive EZH2 was able to bind to its target genes *in vivo* both in the presence and absence of EZH1, if at slightly reduced levels.

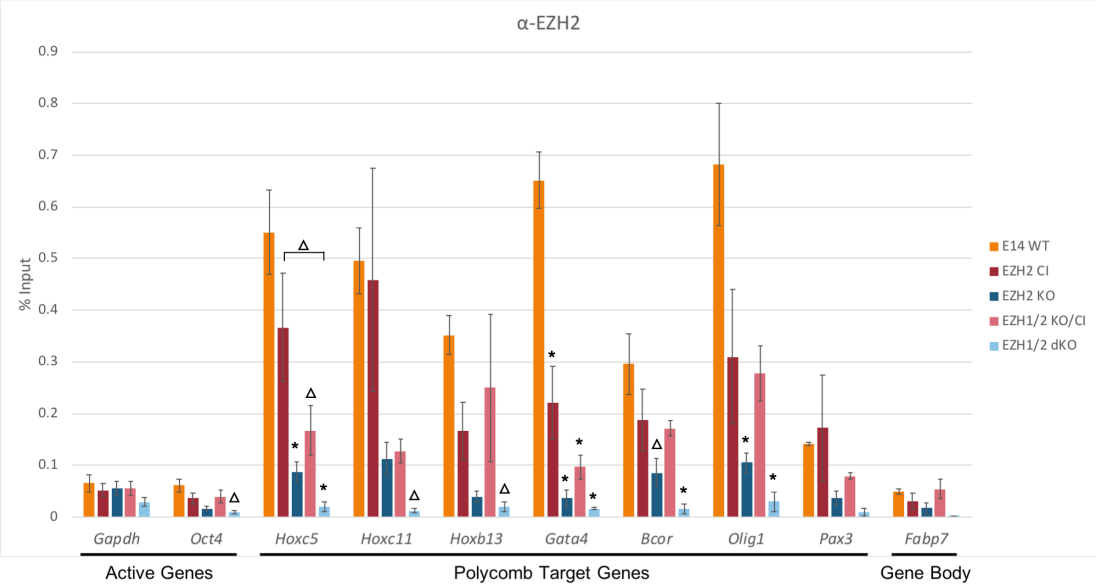


Figure 4.3 EZH2 enrichment in E14 control and PRC2 mutant mESCs. ChIP-qPCR for EZH2 in E14 control (WT), EZH2 Y726A (CI), EZH2 KO, EZH1/2 KO/Y726A (KO/CI) and EZH1/2 dKO cells. Enrichments were normalised to input, with error bars representing standard error of the mean across three biological replicates. * denotes a significant difference to WT with a p value of less than 0.01, and Δ a p value of less than 0.05. Significant differences between the mutants are indicated using brackets.

4.2.3 Analysis of H3K4me3 levels at bivalent genes

I next aimed to determine the consequences of the altered EZH2 recruitment and H3K27me3 deposition in the PRC2 mutant cell lines on H3K4me3. This histone modification is found at the vast majority of Polycomb target genes in mESCs, resulting in a bivalent chromatin state. As expected in the E14 control cells, H3K4me3 was most highly enriched at active genes (Fig. 4.4). Lower, but still reliably detectable levels of this mark were found at the Polycomb target genes with the lowest enrichments of H3K4me3 found at *Hoxc11* and *Hoxb13*. As expected, there was no enrichment of this mark detected at the gene body of *Fabp7*.

In the EZH2 single mutants levels of H3K4me3 at the active genes were largely unchanged except for a slight increase at *Gapdh* in the EZH2 KO cells. At the Polycomb target genes the enrichment of H3K4me3 was similar in the EZH2 Y726A cells to the E14 control, with a slight increase at some genes such as *Gata4* and *Hoxb13*. In the EZH2 KO cells there was an increase of the enrichment of H3K4me3 found at all Polycomb target genes apart from *Hoxc11*. This increase was found to be more pronounced than that seen in EZH2 Y726A cells, apart from at *Hoxb13* and *Gata4*. Levels of H3K4me3 detected at the *Fabp7* gene body remained negligible. Whereas the levels of H3K4me3 at active genes remained unchanged in the EZH1/2 double mutants, the enrichment of this histone modification was increased at the three Hox genes and *Gata4* for both EZH1/2 dKO and EZH1/2 KO/Y726A, with the biggest increases consistently found for the EZH1/2 KO/Y726A. At the other 3 Polycomb bound genes assayed there was also found to be an increase of H3K4me3 in both double mutants but this increase was more modest for the EZH1/2 dKO cells at these genes. As with the EZH2 single mutants, there was no increase in the level of H3K4me3 detected at the *Fabp7* gene body.

Overall I found that the EZH2 single mutants only underwent a small increase in H3K4me3 deposition at Polycomb target genes, the least of which was detected in the EZH2 Y726A with a slightly larger increase found in the EZH2 KO cells. The EZH1/2 double mutants had a much stronger effect on levels of H3K4me3, particularly at the Polycomb target genes which had the lowest enrichments in the E14 control cells. While both double mutants induced an increase in this histone modification at all the Polycomb target genes in this assay, the highest enrichments were consistently found in the EZH1/2 KO/Y726A.

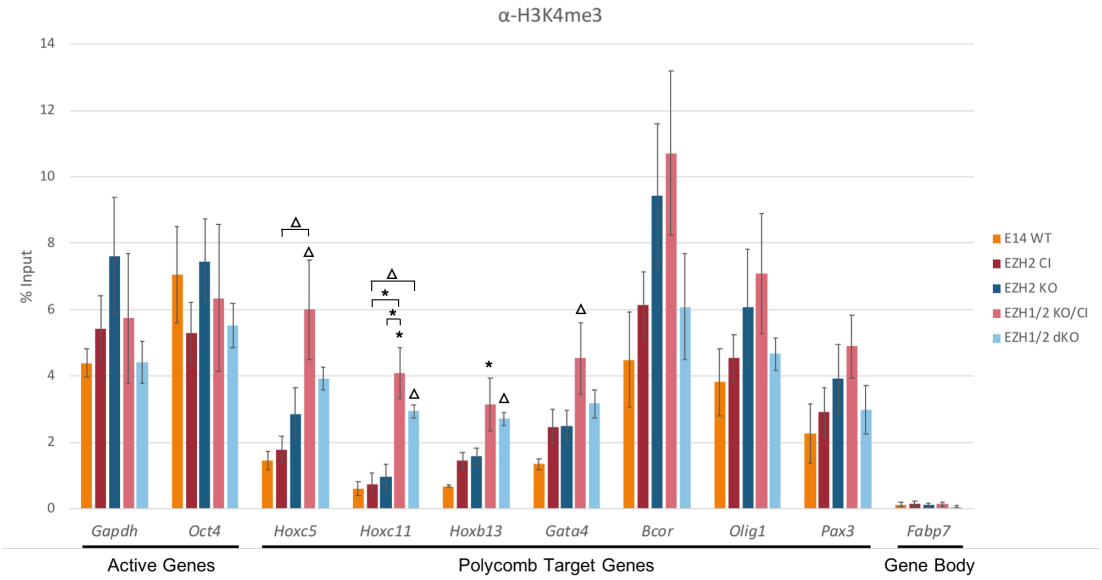


Figure 4.4 H3K4me3 enrichment in E14 control and PRC2 mutant mESCs. ChIP-qPCR for H3K4me3 in E14 control (WT), EZH2 Y726A (CI), EZH2 KO, EZH1/2 KO/Y726A (KO/CI) and EZH1/2 dKO cells. Enrichments were normalised to input, with error bars representing standard error of the mean across three biological replicates. * denotes a significant difference to WT with a p value of less than 0.01, and Δ a p value of less than 0.05. Significant differences between the mutants are indicated using brackets.

ChIP-seq experiments were additionally carried out for EZH2, H3K27me3 and H3K4me3 in E14 control cells and all the PRC2 mutant cell lines generated in this thesis, however time constraints have not allowed for the full analysis of the data at this time. As such, I was not able to include any of the data from these experiments as of August 31st 2019.

4.2.4 Analysis of RING1B binding to bivalent genes

To assess the binding of PRC1 to the Polycomb target genes in the PRC2 mutant cell lines, I performed ChIP-qPCR using an antibody against RING1B, a subunit of the complex that is found both in canonical and non-canonical PRC1. In the E14 control there was a strong enrichment of RING1B at the all the Polycomb target genes, while there was no RING1B detected at either the *Fabp7* gene body or the active genes (Fig. 4.5). In the EZH2 single mutants, RING1B enrichment was reduced to a similar extent at all the Polycomb target genes in both EZH2 Y726A and EZH2 KO. There was however a large variation in the reduction of RING1B by gene, from more than half lost at *Hoxb13* and *Gata4* to around a

third lost at *Hoxc5* and *Pax3*. This variation did not seem to correlate with the variations in loss of H3K27me3 or the levels of EZH2 binding in these cell lines. In both EZH1/2 double mutants the Polycomb target genes were completely depleted for RING1B with the levels detected close to the background levels found at the active genes and the gene body of *Fabp7* and very little difference between the two double mutants. This depletion of RING1B in these cell lines closely mirrors that seen in the H3K27me3 ChIP. This experiment showed that the loss of EZH2 alone had a modest effect on the binding of RING1B to its target genes. This effect was greatly increased with the additional knockout of EZH1 in which case the enrichment of RING1B at these genes was completely lost. It appeared that the presence of a catalytically inactive form of PRC2 at these genes did not mitigate this effect at all indicating RING1B, and thus PRC1 binding to these sites must rely heavily on H3K27me3.

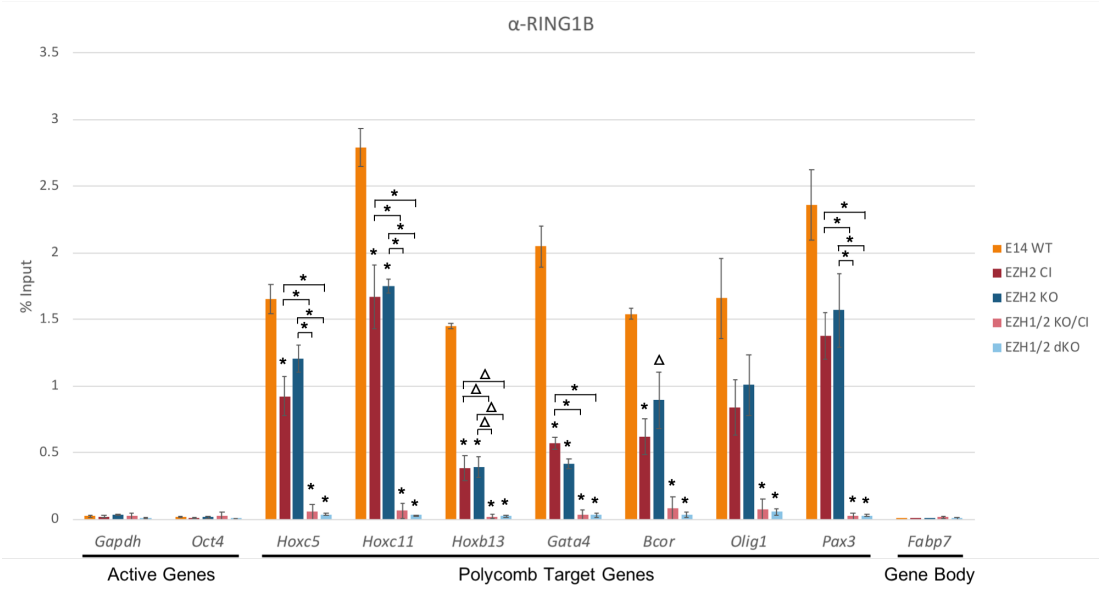


Figure 4.5 RING1B enrichment in E14 control and PRC2 mutant mESCs. ChIP-qPCR for RING1B in E14 control (WT), EZH2 Y726A (CI), EZH2 KO, EZH1/2 KO/Y726A (KO/CI), and EZH1/2 dKO cells. Enrichments were normalised to input, with error bars representing standard error of the mean across three biological replicates. * denotes a significant difference to WT with a p value of less than 0.01, and Δ a p value of less than 0.05. Significant differences between the mutants are indicated using brackets.

4.3 Conclusions

Overall the experiments in this chapter gave me a more detailed picture of how PRC1 and PRC2 behave at Polycomb target genes and how bivalency is maintained at these genes in mESCs. Although I found that H3K27me3 was strongly reduced in the EZH2 single mutant cells, it was still present at some Polycomb targets with varying abundance. This residual H3K27me3 was placed by EZH1 and the reduced levels detected highlighted the inability of EZH1 to compensate for the absence of EZH2 due to the reduced catalytic activity of this subunit as compared to EZH2. The lack of H3K27me3 detected in the EZh2 single mutants at the *Fabp7* gene body may be reflective of the lower overall catalytic activity of EZH1 as compared to EZH2, but also the much lower sensitivity of this subunit to allosteric activation mediated by Jarid2 and EED, which for the EZH2 subunit allows for spreading of H3K27me3 from the initial site of deposition (Margueron et al., 2008; Margueron et al., 2009; Son et al., 2013; Lee et al., 2018; Oksuz et al., 2018). The complete loss of H3K27me3 that was observed in the EZH1/2 double mutants further confirmed that the residual H3K27me3 detected in the EZH2 single mutants is the work of EZH1 as the additional loss of this abolished all signal for H3K27me3.

The levels of EZH2 binding at Polycomb target genes detected in the EZH2 Y726A and EZH1/2 KO/Y726A showed that, although H3K27me3 did appear to increase the affinity of EZH2 for its targets, EZH2-PRC2 was still able to bind to chromatin even in the total absence of the histone mark. This has recently been confirmed by another study in which the same residue is mutated in the set domain of EZH2 as in our EZH2 Y726A (Lavarone et al., 2019).

Interestingly although both the EZH2 KO and EZH1/2 dKO cell lines had undetectable levels of EZH2 by Western blot there was a difference in the levels of EZH2 detected by ChIP in these cell lines. Although it was much reduced as compared to the E14 control, the binding of EZH2 found in the EZH2 single KO was higher than that found in the EZH1/2 dKO in almost every gene assayed. It is possible that there may be some cross-reactivity between EZH2 and EZH1 for this antibody that could explain the low residual signal detected in the EZH2 single KO. The peptide used to produce the antibody used in this experiment was stated as being a fragment around Arg354 of EZH2. Although there are many differences in the protein sequence between EZH1 and EZH2 around this amino acid it is not the most divergent part of the protein and so does not exclude the

possibility of some cross-reactivity between the two paralogues.

I found that in all PRC2 mutants assayed in this chapter, there was an increase in the abundance of H3K4me3 at Polycomb target genes irrespective of the level of H3K4me3 at these genes in the E14 control. The increase in H3K4me3 was found to be much more substantial in the EZH1/2 double mutant cells than the EZH2 single mutants, indicating that residual H3K27me3 at the Polycomb target genes is required for the inhibition of the deposition of tri-methylation at H3K4. Interestingly the increase in H3K4me3 seen in the EZH1/2 double mutants seems to be stronger in the EZH1/2 KO/Y726A even though PRC2 is still bound to these genes. This might indicate a stimulatory effect on the deposition of H3K4me3 by the presence of PRC2. This effect could contribute to the establishment of bivalent domains, allowing H3K4me3 to be deposited on chromatin that carries a repressive signature.

The low enrichments of RING1B at Polycomb genes in the EZH1/2 double mutants was surprising given that it is known to be able to be recruited independently of PRC2 as part of variant PRC1 complexes. It appears that at the Polycomb bound genes assayed in these experiments, the main mode of recruitment for PRC1 is via its binding to the H3K27me3 deposited by PRC2, indicating that canonical PRC1 represents the dominant species of PRC1 at these sites.

Chapter 5

Role of Polycomb in the 3D organisation of the *Lhx5* gene

5.1 Introduction and aims

Polycomb repressive complexes have been shown in many studies to be involved in long range chromatin contacts. Polycomb targeted genes are known to cluster together in the nucleus, both in *Drosophila* (Bantignies et al., 2003; Bantignies et al., 2011; Tolhuis et al., 2011), and in mammalian ESCs (Denholtz et al., 2013; Vieux-Rochas et al., 2015), with a higher enrichment of H3K27me3 increasing the frequency of long-range contacts detected (Vieux-Rochas et al., 2015). These contacts between Polycomb target genes were found to be dependent on the PRC2 complex in mESCs, as the knockout of the EED subunit led to their dissociation (Denholtz et al., 2013; Joshi et al., 2015). This depletion of the PRC2 complex did not affect TAD structure in these cells suggesting that Polycomb mediated contacts and TADs are established independently from each other (Nora et al., 2012). In embryonic carcinoma cells, the *Gata4* gene, a Polycomb target, was found to have interactions with several distal regions through a multi-loop chromatin structure that kept this gene in a transcriptionally inactive state, and this chromatin structure was dependent on the presence of EZH2 (Tiwari et al., 2008b). It was recently found that PRC2 is involved in additional chromatin looping events that mediate the contact of enhancers and promoters of poised genes, and are lost in the absence of the EED subunit (Rada-Iglesias et al., 2011; Cruz-Molina et al., 2017). This looping occurs in ESCs at a stage when

these genes are not yet expressed. Upon differentiation of these cells towards the neural lineage, it was found that the PRC2-dependent contact of enhancers and promoters in ESCs was required for the timely expression of certain genes necessary for the transition of pluripotent cells into neural progenitors.

It has been shown *in vitro* that PRC2 is able to form a dimeric structure by associating with itself (Wu et al., 2013; Davidovich et al., 2014). These structures have been found to be able to take the form of homodimers containing either 2 copies of EZH1 or of EZH2, or heterodimers containing one copy each of both EZH1 and EZH2. It has been suggested that this dimeric state may be involved in recruiting PRC2 to its targets as well as serving as an additional layer of regulation that would depend on the relative levels of EZH1 and EZH2 in cells. It is also conceivable that this dimerisation may play a role in mediating the PRC2 dependent contacts formed at Polycomb domains.

Although the existence of PRC2-dependent contacts has now been demonstrated in several studies, it remains unclear whether they are directly mediated by PRC2 itself. As PRC1 has been shown to mediate the majority of local and long-range contacts between Polycomb targets (Shao et al., 1999; Francis et al., 2004; Eskeland et al., 2010; Kondo et al., 2014; Schoenfelder et al., 2015), it is likely that this is also the case at these sites that were found to be PRC2 dependent. In this case, the essential role played by PRC2 in the establishment of these contacts would be to enable the recruitment of PRC1.

I aimed to determine what effect the PRC2 mutants generated for this thesis might have on the contacts in which PRC2 is involved. I hypothesised that the catalytically inactive EZH2 may be able to retain some of PRC2's roles in mediating chromatin contacts, perhaps through its dimerisation or interaction with other proteins, which would then be lost when this protein is knocked out. To address this question, I probed the state of contacts at a gene that is known to be a Polycomb target. For this I made use of all of the PRC2 mutant cell lines as well as two PRC1 mutant cell lines and an EED KO cell line. The approach I used in this chapter was fluorescence *in situ* hybridisation, or FISH. This method allows for the marking of two or more loci in a cell by hybridisation with fluorescently labelled probes, which in turn allows the distance between the loci to be measured. Through this technique it is possible to determine whether two loci are co-localising or not.

To assess the role played by PRC2 in mediating chromatin contacts I decided to

focus on a contact that has already been shown to depend on the presence of PRC2. For this purpose I chose the *Lhx5* gene. This is a developmental gene involved in differentiation towards the neural lineage and is a bivalent Polycomb target in mESCs. The enhancer of this gene also carries a bivalent signature and in mESCs, is in physical contact with its promoter (Cruz-Molina et al., 2017). The contact of these two elements was shown to be dependent on the PRC2 complex, however this study did not discern between the presence of the complex itself and the presence of the trimethylated H3K27.

By comparing the distances between the enhancer and promoter of this gene in the E14 control cells against those measured in the PRC2 mutant as well as PRC1 mutant cells, I sought to begin elucidating the mechanism by which the PRC2 complex mediates these chromatin contacts.

The ChIP-qPCR and FISH experiments in this chapter were performed with the help of Giulia Bartolomucci under my direction with advice from Katy McLaughlin.

5.2 Generation and initial characterisation of mESCs expressing a H3K27me3-binding deficient CBX7

In addition to the PRC2 mutant cell lines described in the preceding chapters, I also used two PRC1 mutant cell lines for the experiments in this chapter. The RING1B KO cells were generated in the Bickmore lab in 2015 (Illingworth et al., 2015). The CBX7 F11A mutant was generated in the Voigt lab using CRISPR Cas9 (see Fig. 5.1 for CRISPR targeting strategy used). The mutations incorporated in this cell line aimed to block the ability of this protein to bind to the H3K27me3 mark deposited by PRC2.

This CBX subunit was chosen in particular as it has been shown to be the most highly expressed of all CBX proteins in mESCs and is entirely dependent on H3K27me3 for its recruitment to chromatin (Morey et al., 2012). The phenylalanine 11 residue resides within the highly conserved chromodomain of CBX7 and its mutation to an alanine has been shown to disrupt its binding to chromatin (Li et al., 2010).

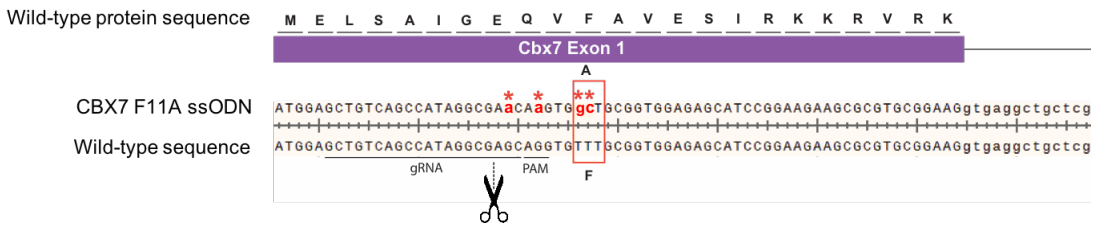


Figure 5.1 Design of CBX7 F11A CRISPR targeting strategy. Alignment of E14 Wild-type sequence of exon 1 of *Cbx7* gene with the ssODN designed to induce CBX7 F11A mutations (uppercase is exonic sequence, lowercase is intronic). gRNA hybridisation site and PAM sequence are labelled, Cas9 cut site is indicated with scissors. Point mutations in ssODN are indicated by red asterisks (lowercase in red), in frame stop codons are labelled.

Before using these cell lines in the FISH experiments I first sought to assess their effect on the levels of H3K27me3 and PRC1 at Polycomb targets. As can be seen in Fig. 5.2A, there was an enrichment of H3K27me3 detected at Polycomb targets *Hoxc5* and *Fzd1*, as well as at both the enhancer and promoter of the *Lhx5* gene. Neither the RING1B KO nor the mutation of the CBX7 protein altered the enrichment of the H3K27me3 mark at the control Polycomb target genes *Hoxc5* and *Fzd1*, nor was there any change at the enhancer or promoter of *Lhx5*. I next sought to determine how the binding of RING1B was affected at these loci. As shown in Fig. 5.2B, I found that although RING1B was enriched at both *Hoxc5* and *Fzd1*, it was only detected at the promoter of *Lhx5* but not its enhancer in the E14 control cells. This enrichment of RING1B was strongly depleted in both the RING1B KO and CBX7 F11A cell lines. These ChIP experiments showed that although the mutations in both the RING1B KO and the CBX7 F11A cells led to a massive loss of PRC1 binding to the genes assayed, this did not have a great effect on the activity of PRC2 at these sites, as evidenced by the unaffected levels of H3K27me3 detected. These results also demonstrated that H3K27me3 binding by CBX7 appears to be the major determinant of PRC1 recruitment to these sites.

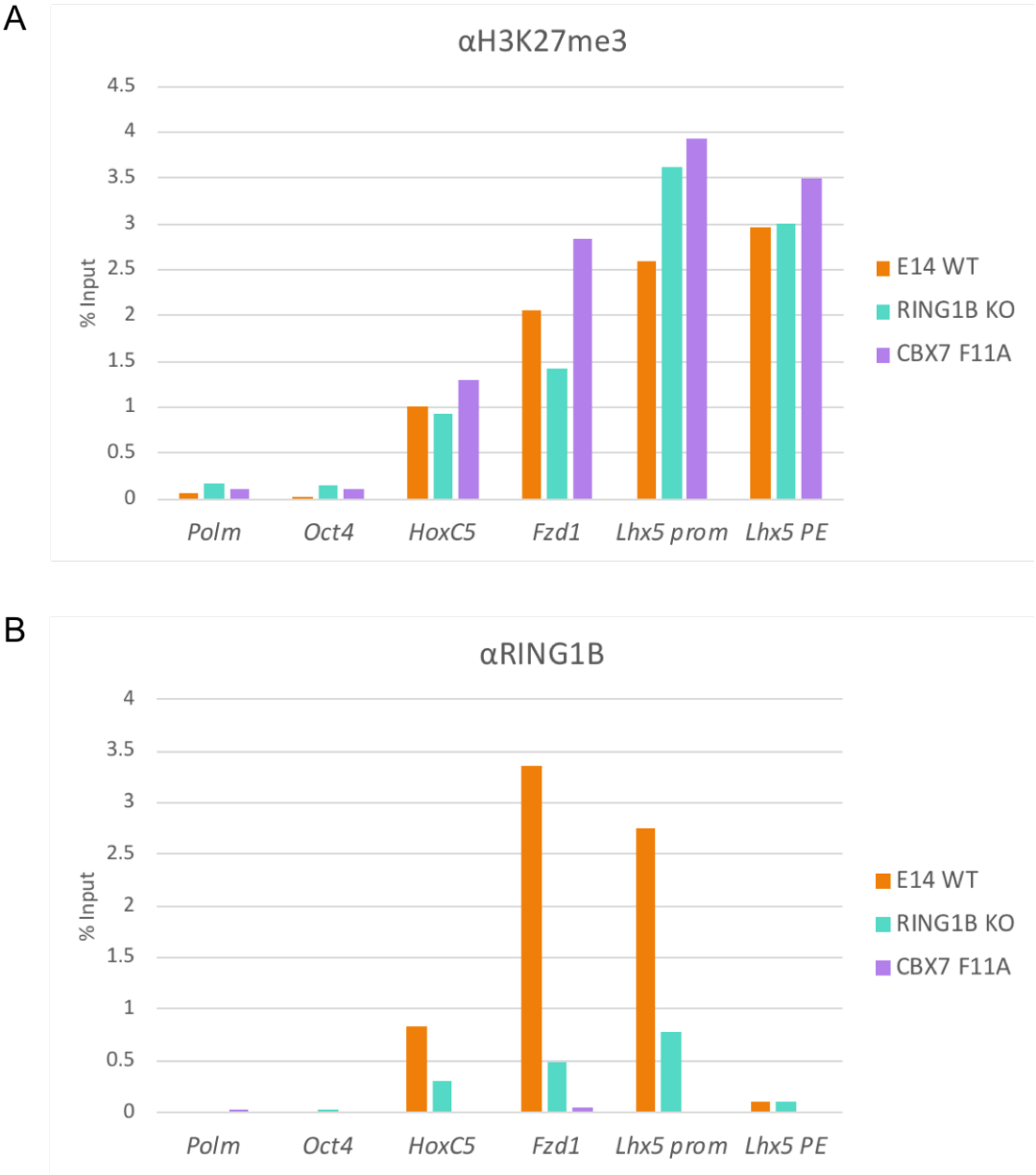


Figure 5.2 H3K27me3 and RING1B enrichments in E14 control and PRC1 mutant mESCs. A. ChIP-qPCR for H3K27me3 in E14 control (E14 WT), RING1B KO and CBX7 F11A cells. **B.** ChIP-qPCR for RING1B in E14 control(E14 WT), RING1B KO and CBX7 F11A cells. Enrichments were normalised to input. Results from one experiment shown.

5.3 FISH experimental design

The fluorescently labelled probes used for the experiments in this chapter were designed to hybridise to the enhancer and the promoter of the *Lhx5* gene. These sites also corresponded to regions that were marked by an enrichment of H3K27me3 as was observed by ChIP-seq (see Fig. 5.3 and Cruz-Molina et al., 2017). A control probe was designed to hybridise to a region downstream of *Lhx5* that was an equal distance from the promoter probe as the enhancer probe. This region has no Polycomb enrichment and therefore its position relative to the promoter, although the same distance from the promoter as the enhancer, was not expected to be affected in the Polycomb mutant cell lines.

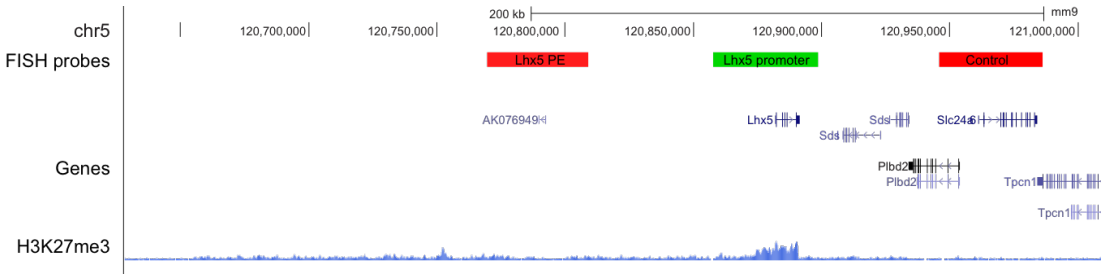


Figure 5.3 Position of probes used in *Lhx5* FISH experiments. Schematic showing the position of probes corresponding to enhancer and promoter of *Lhx5* as well as equidistant control probe, relative to *Lhx5* location. Lower track shows H3K27me3 enrichments across this region (ENCODE)

5.4 Analysis of a PRC2 dependent enhancer-promoter contact - *Lhx5* locus

The results of the FISH analysis are shown in Fig. 5.5. In the EZH2 single mutant cell lines there was not an increased distance between the probes corresponding to the enhancer and the promoter of the *Lhx5* gene as compared to the E14 control cells. For the EZH2 KO cells however there even appeared to be a decrease in the interprobe distance as compared to the control. An additional observation that was made for the EZH2 single mutant cell lines was the broader variation of individual values for the measured interprobe distances.

In the EZH1/2 dKO cells the average interprobe distance was significantly increased as compared to the E14 control, indicating a reduced propensity for

contacts between the enhancer and promoter of *Lhx5* to form in these cells. A comparable increase in the interprobe distance was also observed in the EED KO cell line (Fig. 5.5A). This increased spacing between the enhancer and promoter of *Lhx5* in these two cell lines further confirmed the findings from the Cruz-Molina 2017 paper that this contact at this locus in mESCs is disrupted upon depletion of PRC2.

I next assessed the enhancer-promoter contact at this same locus in two PRC1 mutant cell lines. In the RING1B KO there was an increase in the interprobe distance as compared to the E14 control cells (although just below the threshold for significance), suggesting that the PRC1 complex was also involved in establishing the contact between these elements. An increase in the interprobe distance was also observed in the CBX7 F11A cells (Fig. 5.5A). This result showed that H3K27me3 was essential for the recruitment of canonical PRC1 to this locus and the formation of the enhancer-promoter contact.

There was no significant difference observed in the distance between the promoter and control probes in any of the cell lines analysed as compared to the E14 control (Fig. 5.5B). Overall these experiments suggest that the contact between the poised enhancer and promoter of the *Lhx5* gene required the contribution of both the PRC2 and PRC1 complexes.

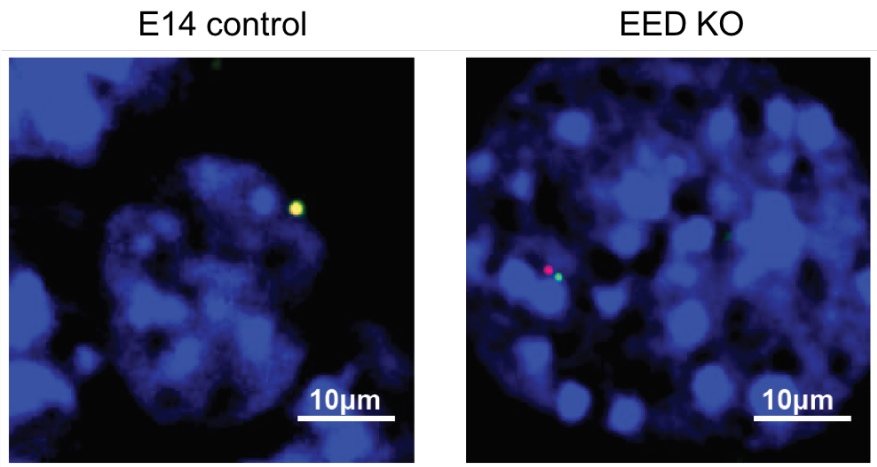


Figure 5.4 *Lhx5* enhancer and promoter probe hybridisation in mESCs. Representative images of probe hybridisation signal (shown in green and red) for the *Lhx5* locus in E14 control and EED KO cells. DAPI counterstain is shown in blue.

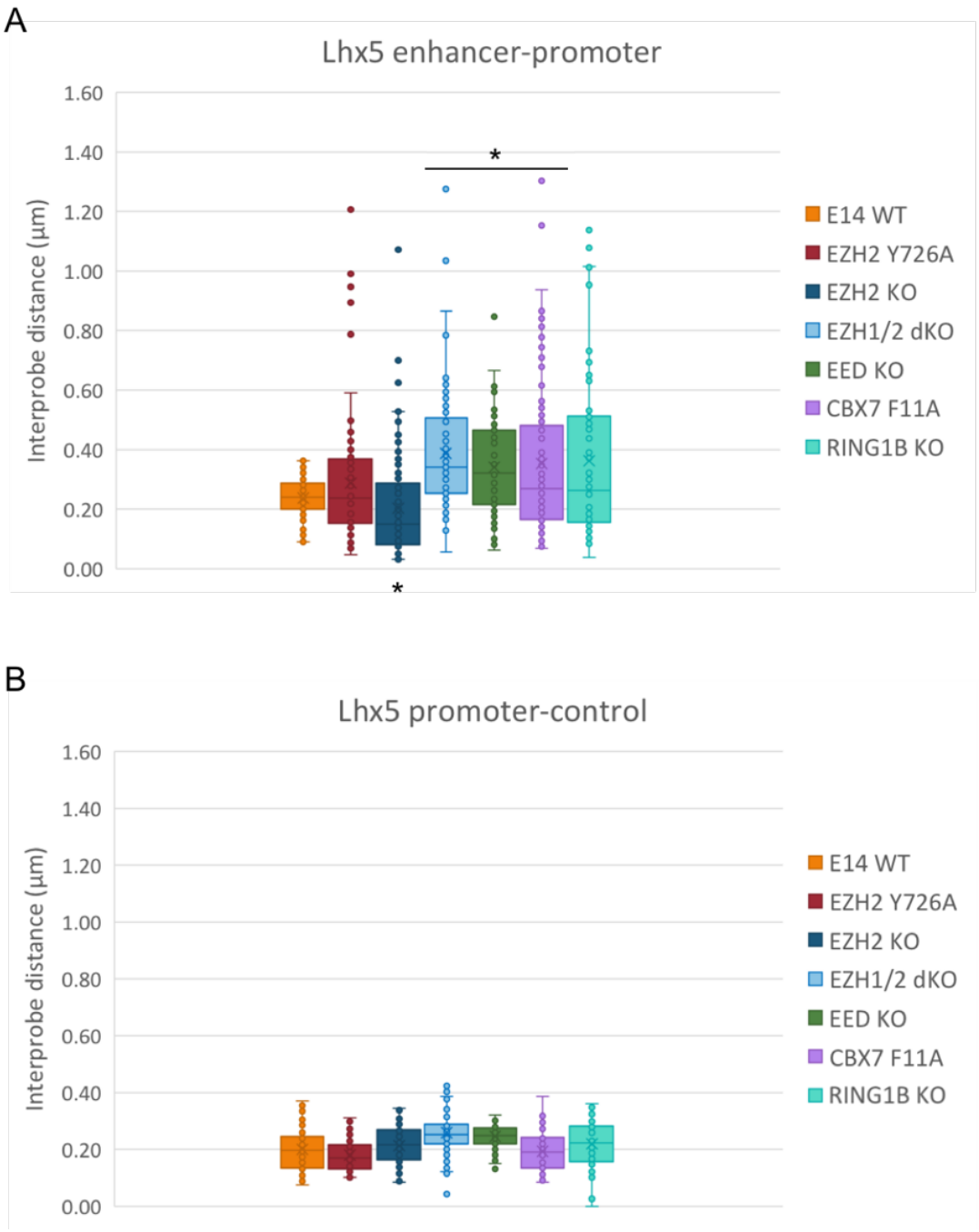


Figure 5.5 FISH analysis of *Lhx5* enhancer-promoter contact in E14 control and PRC1 and PRC2 mutant mESCs. **A.** Boxplot showing the distribution of distances between *Lhx5* promoter and *Lhx5* enhancer probe hybridisation signals in E14 control, EZH2 Y726A, EZH2 KO, EZH1/2 dKO, EED KO, CBX7 F11A and RING1B KO cell lines. Significant differences (p value $\leq$0.05) as determined by a Kruskal-Wallis test followed by a Mann-Whitney test, are marked by an asterisk. **B.** Boxplot showing average distance measured between probes corresponding to *Lhx5* promoter and control locus in E14 control (E14 WT), EZH2 Y726A, EZH2 KO, EZH1/2 dKO, CBX7 F11A and RING1B KO cell lines.

5.5 *Lhx5* expression in mESCs

As will be described in detail in chapter 6, I performed RNA-seq on both the E14 control and the EZH2 single and EZH1/2 double mutant cell lines. I used this data to further investigate the state of the *Lhx5* locus. As can be seen in Fig. 5.6, very little transcription was detected across the *Lhx5* gene in the E14 control cells, as is expected of a gene involved in neuronal development in this undifferentiated cell type. In the EZH2 Y726A cells the levels of transcription appeared similar to those detected in the E14 control cells, while in the EZH2 KO cells, there seemed to be a slight decrease in the level of transcripts detected across *Lhx5*. For both of the EZH1/2 double mutant cell lines there was a marked increase in transcription of this gene, with this increase being marginally larger in the EZH1/2 dKO cells, indicating derepression of the gene in complete absence of PRC2 activity.

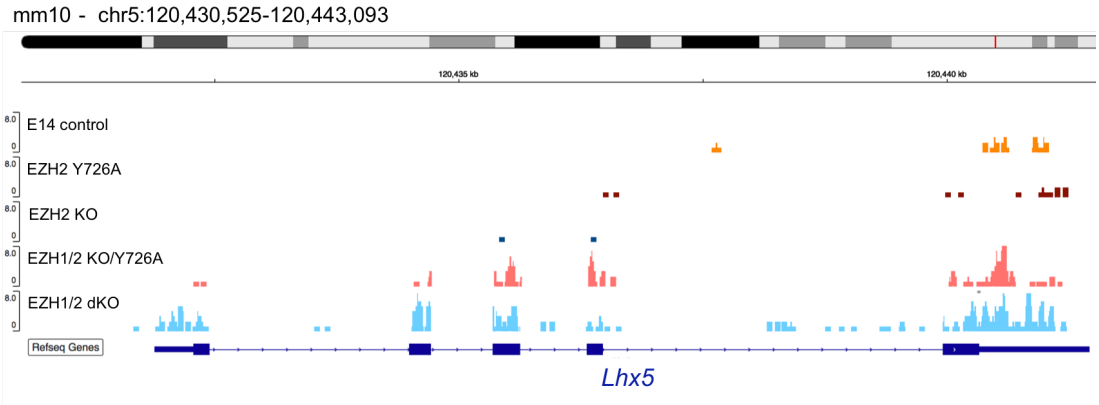


Figure 5.6 RNA-seq reads detected across the *Lhx5* gene in E14 control and PRC2 mutant cell lines. Screenshot of the IGV genome browser tool. RNA-seq tracks for E14 control (E14 WT), EZH2 Y726A, EZH2 KO, EZH1/2 KO/Y726A and EZH1/2 dKO were aligned with the mm10 version of the mouse genome assembly to display reads detected for the *Lhx5* gene. Tracks shown are representative of one of three replicates.

5.6 Conclusions

Taken together, the preliminary results in this chapter indicate that the enhancer-promoter contact at the *Lhx5* gene, although dependent on the presence of PRC2, appears to require this complex in order to recruit the PRC1 complex which is likely to be the direct mediator of compaction. The loss of PRC2 from cells in the EZH1/2 dKO and EED KO backgrounds, as well as the loss of PRC1 binding seen in the RING1B KO and CBX7 F11A cells, both led to an increase in the distance between the *Lhx5* enhancer and promoter probes, demonstrating that both complexes were required for the maintenance of this contact. Furthermore, the depletion of PRC2 in the EZH1/2 dKO cells led to a loss of PRC1 binding to Polycomb targets (see chapter 4), but the depletion of PRC1 from its targets in the RING1B KO and CBX7 F11A cells did not result in a loss of H3K27me3 from these sites. We can presume that PRC2 was also maintained at these loci in the absence of PRC1, but further ChIP experiments probing for EZH2 or other core subunits of PRC2 would be required. We can then conclude from this that, while PRC2 is required for the contact between the *Lhx5* enhancer and promoter, this complex alone cannot directly enable it, but rather acts through the recruitment of PRC1, via the H3K27me3 mark, to this site. PRC1 is then itself able to mediate the contact, likely through the action of its PHC subunit which can enable this type of interaction through the oligomerisation of its SAM domain (Isono et al., 2013; Kundu et al., 2017). This is in accord with the finding of Tiwari et al. that the losses of contacts between two given sites observed following the depletion of EZH2 correlated more strongly with the degree of H3K27me3 depletion than with the loss of EZH2 at these same sites (Tiwari et al., 2008b). The results from the CBX7 F11A cells indicated that CBX7 plays a major role in the recruitment of PRC1 to these sites and further stressed the importance of H3K27me3 in this recruitment. It remains to be determined whether the presence of H3K27me3 is sufficient for the establishment of the contact between the enhancer and promoter of *Lhx5*, or whether the additional presence of the PRC2 complex itself at these sites is also required.

I found that the presence of EZH1 alone appears to be sufficient to maintain the enhancer-promoter contact. This correlates with the results from chapter 4 in which I observed that in the EZH2 single mutant cell lines, the majority of RING1B was maintained at Polycomb genes as opposed to the complete loss that occurred in the EZH1/2 double mutant cells. Together these results further

demonstrate that in mESCs, very low levels of H3K27me3 appear to be sufficient to maintain the majority of PRC2 function.

Surprisingly, in the EZH2 KO cells I found that the interprobe distance was actually decreased as compared to the E14 control. This may be due to the presence of EZH1 which is known to have a chromatin compaction activity that EZH2 does not possess (Margueron et al., 2008). In the EZH2 KO cells, as there is no EZH2-PRC2 present, there may be an increased binding of EZH1 as this site which could then lead to an even more condensed state than was observed in the E14 control. This may not have been observed in the EZH2 Y726A cells as the EZH2-PRC2 species is maintained and bound to Polycomb targets, preventing an increase in the binding of EZH1. As was shown by the RNA-seq data across the *Lhx5* locus, there was a derepression of this gene only in the case of the EZH1/2 double mutant cells. In the EZH2 KO there was a slight decrease in the level of transcription detected, which could be a reflection of the more condensed state that was found at this locus in the FISH experiment. This suggests that EZH2 may play a role in prohibiting an overly compact state of the genes it binds.

As previously stated, the work in this chapter is preliminary and experiments such as the ChIP-qPCR for H3K27me3 and RING1B need to be repeated and the FISH experiments performed in additional clones as well as in the EZH1/2 KO/Y726A cells. In the longer term additional individual contacts will be analysed by FISH and perhaps further assessed by 4C.

Chapter 6

Gene expression in PRC2 mutant mESCs

6.1 Introduction and aims

In previous studies it has been shown that although a knockout of EZH2 results in massive loss of H3K27me3 from all PRC2 targets and depletion of PRC2 binding to chromatin, this does not result in a strong phenotype at the level of gene expression (Chamberlain et al., 2008; Riising et al., 2014; Shen et al., 2008). Although it might be expected that the loss of this repressive chromatin signature at Polycomb target genes would lead to their upregulation, that appears not to be the case for the vast majority of genes. Derepression of these genes has only been found to occur in a situation where the PRC2 complex is completely disrupted, for instance by knockout of the core subunits EED or SUZ12 (Pasini et al., 2007; Chamberlain et al., 2008; Shen et al., 2008). Perhaps reflecting these moderate changes in gene expression, neither EZH2 knockout nor SUZ12 or EED knockout leads to any defects in self-renewal or viability in mESCs (Pasini et al., 2007; Chamberlain et al., 2008; Shen et al., 2008).

Although EZH1 and EZH2 have quite different expression profiles and have differing enzymatic activity towards H3K27 (Bracken, 2003; Margueron et al., 2008), it has been shown that there is a certain redundancy between these two subunits at least in ESCs. As has been shown in previous studies (in agreement with the ChIP-seq experiments from chapter 4) in the absence of EZH2, EZH1 is able to place H3K27me3 at most PRC2 targets albeit at a much lower level and

a narrower distribution across the target sites.

The aim of the work described in this chapter was to obtain a comprehensive view of how the different mutations of PRC2 affect gene expression in mESCs and to determine how these changes relate to Polycomb occupancy and abundance of H3K27me3 and H3K3me3 at these genes by comparison to ChIP-seq data. Although we expected only minor changes in gene expression in the EZH2 single mutants compared to E14 control cells based on previous work, there may be differences between the EZH2 KO and EZH2 Y726A. The greatest changes in expression should be observed in the EZH1/2 double mutants as it has been shown previously that complete loss of PRC2 and H3K27me3 leads to significant upregulation of Polycomb genes (Pasini et al., 2007; Chamberlain et al., 2008; Shen et al., 2008). Here, I was interested to see if the maintenance of a catalytically inactive PRC2 bound to genes could maintain some extent of repression of Polycomb genes, either genome-wide or on a subset of genes affected.

The immunofluorescence staining of cells in this chapter was performed with the help of Katy McLaughlin. The quality control and processing of RNA-seq raw data was performed under my direction by Shaun Webb.

6.1.1 Expression of pluripotency factors

I first aimed to determine the expression patterns of a known pluripotency factor in the PRC2 mutant cell lines to provide some insight into the "stemness" of these cell lines. The intention was to verify that any changes in gene expression observed in the following experiments were not caused by cells exiting pluripotency and exhibiting characteristics of early stages of differentiation. All mutant cell lines were stained for OCT4 as well as expression of TUJ1, a neuron-specific beta-tubulin, often used as a neuronal marker gene and not expected to be expressed in undifferentiated ESCs. As can be seen in Fig. 6.1, there were no striking changes in the expression pattern of *Oct4* in any of the mutants as compared to the E14 control. Similarly, there was no obvious increase in the expression of *Tuj1*. RT-qPCR analysis of the expression of *Oct4* and *Nanog* in all cell lines, as seen in Fig. 6.2, similarly showed that there was no strong downregulation of either of these genes in any of the PRC2 mutants relative to the E14 control. This is consistent with previous studies in which similar mutations of PRC2 did not have a significant effect on the expression of pluripotency factors (Pasini et al., 2007; Shen et al., 2008).

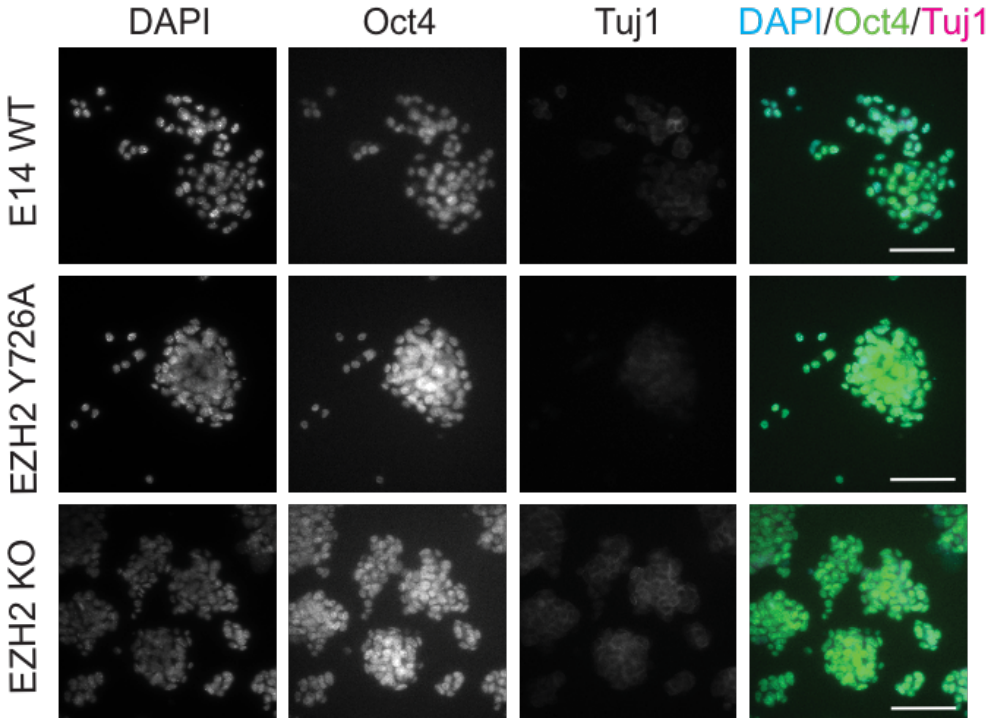


Figure 6.1 Immunofluorescence analysis of *Oct4* and *Tuj1* expression in E14 control, EZH2 single and EZH1/2 double mutant mESCs. DAPI (far left) OCT4 (left) and TUJ1 (right) staining of E14 control, EZH2 Y726A, EZH2 KO, EZH1/2 KO/Y726A and EZH1/2 dKO ESCs. Merge is shown in panels on far right, DAPI in cyan, OCT4 in green and TUJ1 in magenta. Scale bar corresponds to 100 µm.

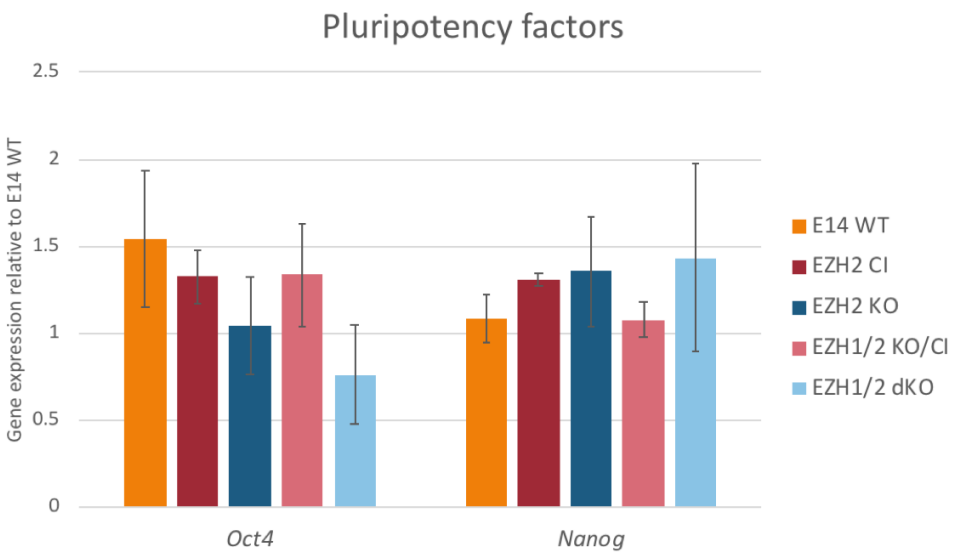


Figure 6.2 Expression of pluripotency factors *Oct4* and *Nanog* in EZH2 single and EZH1/2 double mutants. Bar chart showing expression of *Oct4* and *Nanog* relative to E14 control. Data was generated by RT-qPCR and normalised to expression of *Gapdh*. Plotted are the mean and standard error of the mean (SEM) across 3 biological replicates.

6.1.2 Expression of select Polycomb genes

To assess the expression of a selection of known Polycomb target genes I performed RT-qPCR on cDNA samples from E14 control ESCs and both EZH2 single and EZH1/2 double mutants. As can be seen in Fig. 6.3 for all but one genes tested there was a modest but detectable increase in expression for every mutant cell line. This increase was of a similar amount at each gene for the EZH2 single mutants, with little consistent difference detected between EZH2 KO and EZH2 Y726A. Overall, the increase in expression of the genes analysed was stronger for the EZH1/2 double mutants than for the EZH2 single mutants. This increase was also more variable gene to gene in the double mutants with the foldchange in expression of *Hoxb13* found to be nearly 4 times higher than that of *Meis2*. Interestingly, the EZH1/2 KO/Y726A mutant was found to express all of the genes assayed (with the exception of *HoxB13*) to a much higher level than the EZH1/2 dKO which displayed a gene expression pattern much closer to that of the EZH2 single mutants. Overall, these results showed that the EZH1/2 double mutants exhibited stronger derepression of the Polycomb target genes tested than the EZH2 single mutants, with the EZH1/2 KO/Y726A mutant appearing to induce the strongest upregulation of these genes. Interestingly however, I found

that not all Polycomb target genes were derepressed in the double mutants, as silencing of *Pax6* was maintained in all cell lines tested.

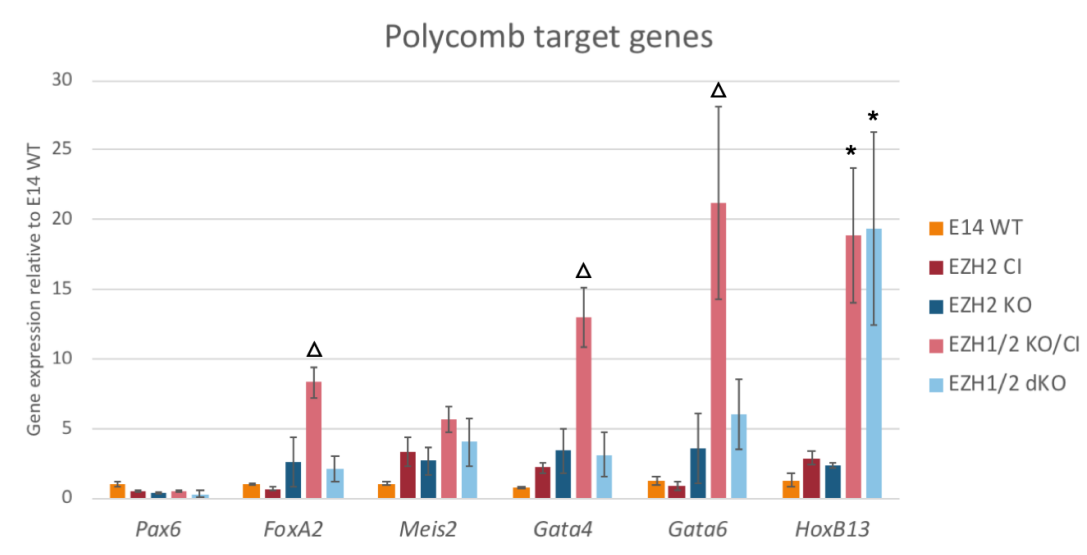


Figure 6.3 Expression of Polycomb target genes in EZH2 single and EZH1/2 double mutants. Bar chart showing expression of known Polycomb target genes in EZH2 Y726A (CI), EZH2 KO, EZH1/2 KO/Y726A (KO/CI) and EZH1/2 dKO cells relative to E14 control. Data was generated by RT-qPCR and normalised to expression of *Gapdh*. Plotted are the mean and SEM across 3 biological replicates. * denotes a significant difference to WT with a p value of less than 0.01, and Δ a p value of less than 0.05.

6.1.3 RNA-seq analysis of gene expression in Polycomb mutants

After having established that the EZH2 and EZH1 mutation affect expression for Polycomb target genes, I performed RNA-seq to assess genome-wide effects on gene expression of the EZH1 and EZH2 mutations. I chose to use three independent clones of each genotype as replicates for these experiments in order to account for any clonal variation in gene expression. Before I was able to move forward with the analysis, a rigorous quality control was performed on the raw data. All samples were found to have between 20 and 40 million unique reads and were tested for, among others, GC bias, overall base composition, read length, and adaptor contamination. Once each sample was confirmed to be of acceptable quality based on each of these criteria, the reads were mapped to the mouse genome. For all samples the percentage of uniquely mapped reads was between 60 and 85% and these were again, screened for the same set of criteria as the raw

data. With the quality control complete and all samples passing the threshold that was set, I was confident in moving onto the analysis of gene expression in each cell line.

6.1.4 Global changes in gene expression

As can be seen in Fig. 6.4 and 6.5, the trends in gene expression changes between E14 control and EZH1/2 mutant ESCs that I observed by RT-qPCR were reflected in the gene expression patterns genome wide. Very few genes were differentially expressed in either of the EZH2 single mutant cells, with approximately 0.3% of all protein coding genes found to be significantly up- or downregulated as compared to the E14 control. Of these 70 or so differentially expressed genes in the EZH2 single mutants, surprisingly, the majority were downregulated relative to the E14 control. In the EZH1/2 double mutants, as expected, a much higher proportion of genes was differentially expressed. Expression of around 10% of all protein coding genes was misregulated in EZH1/2 KO/Y726A cells, slightly fewer than in the EZH1/2 dKO cells in which this rose to 13%. In EZH1/2 KO/Y726A cells a greater number of genes was upregulated than downregulated. This was also true for the EZH1/2 dKO cell lines, although in these cells there were around 300 more genes misregulated than in the EZH1/2 KO/Y726A. In addition, I found that EZH1/2 dKO cell lines had nearly 1.5 times as many genes upregulated by more than a log(2) foldchange of 1 (i.e. upregulated by more than 2 fold) than the EZH1/2 KO/Y726A.

Figure 6.4 Differential gene expression in EZH2 single mutant mESCs. Pie charts representing expression of all protein coding genes in EZH2 Y726A and EZH2 KO cell lines relative to E14 control. Genes significantly upregulated are shown in shades of yellow, genes significantly downregulated are shown in shade of blue. Darker shades correspond to genes differentially expressed by more than a log(2) foldchange of 1 or -1 over E14 control. For this and all subsequent figures presenting RNA-seq data, data is derived from analysis of three independent cell lines per genotype.

Figure 6.5 Differential gene expression in EZH1/2 double mutant mESCs. Pie charts representing expression of all protein coding genes in EZH1/2 KO/Y726A and EZH1/2 dKO cell lines relative to E14 control. Genes significantly upregulated are shown in shades of yellow, genes significantly downregulated are shown in shade of blue. Darker shades correspond to genes differentially expressed by more than a log(2) foldchange of 1 or -1 over E14 control.

6.1.5 Gene expression changes in EZH2 Y726A and EZH2 KO cell lines

As was shown in Fig. 6.4, both EZH2 Y726A and EZH2 KO cell lines misregulated a similar number of genes as compared to the E14 control. I found however, that of these genes, only around half were common to both cell lines (Fig. 6.7A). These 35 genes were found to be expressed in a very similar manner in both EZH2 KO

and EZH2 Y726A cell lines, being either up- or downregulated by approximately the same factor in each mutant cell line relative to the E14 control. For the genes that were found to only be significantly differentially expressed in one or the other cell line, changes in gene expression still followed the same trend and were comparable between the two mutants when this was plotted according to log(2) foldchange relative to the E14 control, but only satisfying the criteria for significance in one but not the other genetic background (Fig. 6.6).

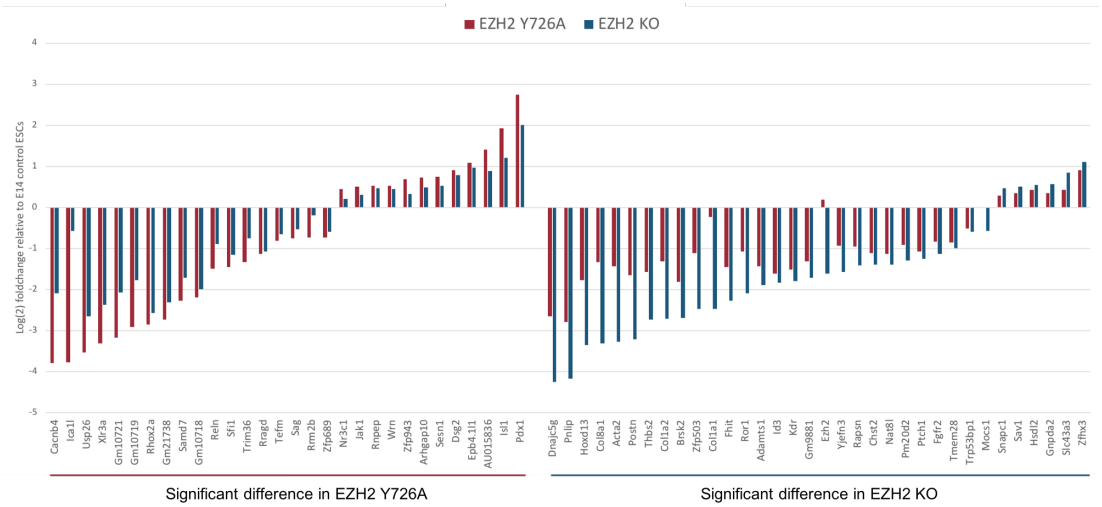


Figure 6.6 Comparison of differentially expressed genes in EZH2 single mutant mESCs. Log(2) foldchange in expression over E14 control in both EZH2 single mutant cell lines of genes for which the threshold of significance was only reached in one of the mutants.

To shed some light on the function of the genes that were found to be differentially expressed in the mutant cell lines I performed gene ontology (GO) analysis (Fig. 6.7B). In line with only a small number of genes being significantly upregulated in the EZH2 single mutant cell lines, very few GO terms were found to be significantly enriched amongst these genes. Only the Sgk3 homodimer complex was found to be enriched amongst the genes upregulated in both single mutants. This is a protein kinase activated by phosphorylation by 3-phosphoinositide-dependent protein kinase-1 (PDK1) (Kobayashi et al., 1999), and has been associated with cell proliferation and cancer cell survival (Basnet et al., 2018). However it should be noted that the tumour suppressor gene *Cdkn2a* which is associated with senescence, was also upregulated in both cell lines as has been previously described both in EED KO and SUZ12 KO cell lines (Bracken et al., 2007; Shen et al., 2008; Lowe and Sherr, 2003).

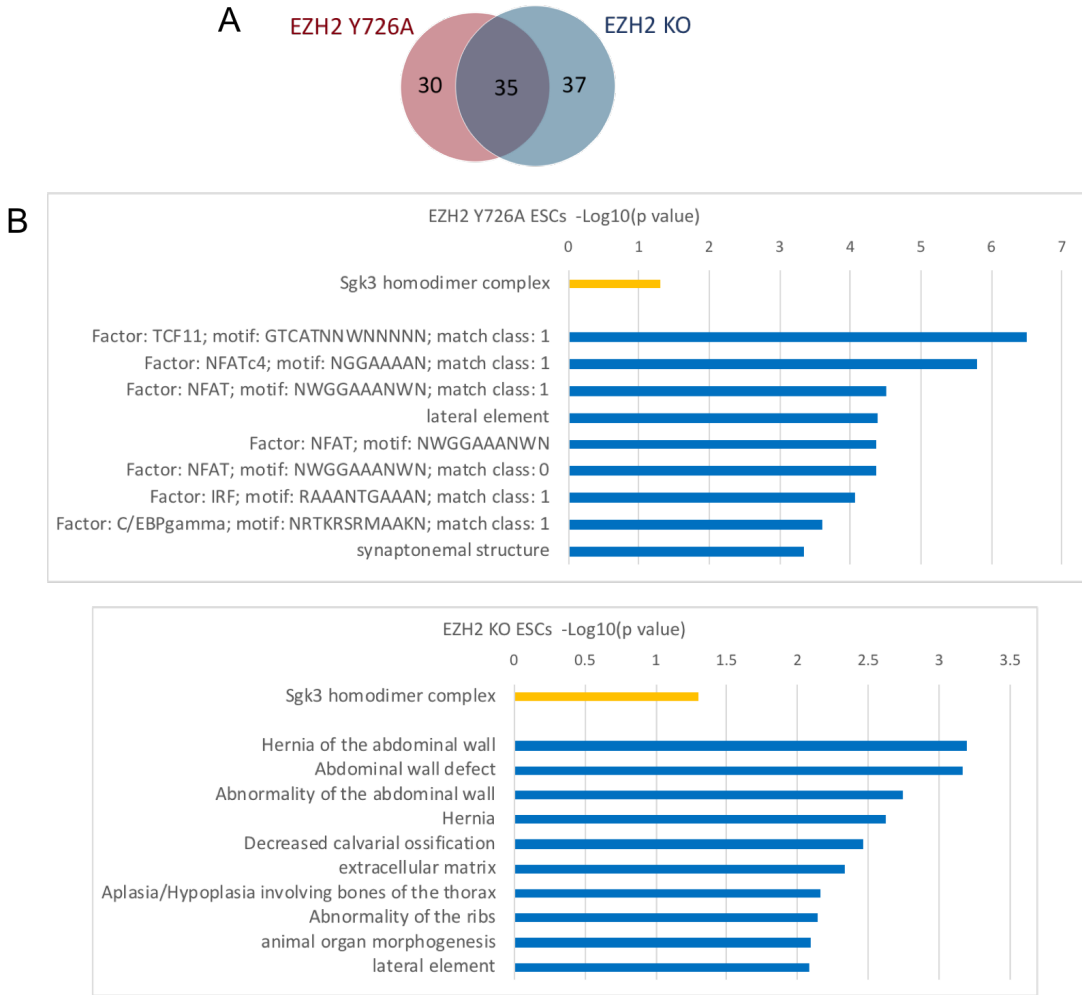


Figure 6.7 Gene ontology of differentially expressed genes in EZH2 single mutant mESCs. A. Venn diagram of genes differentially expressed in EZH2 Y726A and EZH2 KO cells. **B.** GO terms enriched in genes significantly upregulated (yellow) and downregulated (blue) in EZH2 Y726A and EZH2 KO mESCs relative to E14 control.

Amongst the genes that were found to be downregulated there was little overlap between GO terms enriched for EZH2 Y726A and EZH2 KO cell lines. In the EZH2 Y726A cell line there was an enrichment of genes with binding motifs for a few different transcription factors (TF), which were not significantly enriched in the EZH2 KO. Both EZH2 single mutants did however significantly downregulate genes associated with the lateral element and synaptonemal complex, a protein structure present between sister chromatids during prophase of meiosis (Heyting, 1996). The EZH2 KO was found to additionally downregulate genes associated with morphological abnormalities and defects in growth of certain body parts such as abnormality of the ribs and of the abdominal wall. Genes which were

common to all of these GO terms to do with body abnormalities were *Col1a1* and *Col1a2*, both encoding alpha chains of type I collagen. Taken together, these results did not suggest a great difference in the effects on gene expression between the catalytically inactive form of EZH2 and the EZH2 KO.

6.1.6 Gene expression changes in EZH1/2 KO/Y726A and EZH1/2 dKO cell lines

Similarly to what I observed in the EZH2 single mutant cell lines, I found that a considerable number of the differentially expressed genes in the EZH1/2 double mutant cell lines were shared between both genotypes. In the EZH1/2 KO/Y726A cells about two thirds of all differentially expressed genes were also affected in EZH1/2 dKO (Fig. 6.11A). In the EZH1/2 dKO cells, however, more than half of the genes differentially expressed were not significantly affected in the EZH1/2 KO/Y726A.

As with the EZH2 single mutant cell lines, I performed GO term analysis to gain some insight into the function of the deferentially expressed genes (Fig. 6.8 and Fig. 6.9). Analysis of the transcription factor motifs enriched amongst the differentially expressed genes was also performed to provide further information on their regulation, for instance to clarify whether any transcription factors that are direct PRC2 targets could be responsible for the changes in expression of a significant portion of the genes misregulated. This analysis could also be helpful in identifying common traits in the regulatory regions of the genes affected, and possibly shed light on major regulatory pathways influencing the expression of these genes.

In both double mutant backgrounds, there was a strong enrichment in the upregulated genes for terms related to embryo development, body patterning and cell fate commitment, all terms that are also strongly enriched amongst known Polycomb target genes. In addition, I found that in the EZH1/2 dKO many of the genes upregulated contained TF binding motifs that were often quite CG rich (Fig. 6.10), a well-known hallmark of Polycomb bound genes. As for the downregulated genes, there were fewer enriched GO terms identified in EZH1/2 KO/Y726A than in EZH1/2 dKO cells (Fig. 6.8 and Fig. 6.9). No specific biological processes were found to be downregulated in the EZH1/2 KO/Y726A, and only a few TF binding motifs that had little sequence homology were

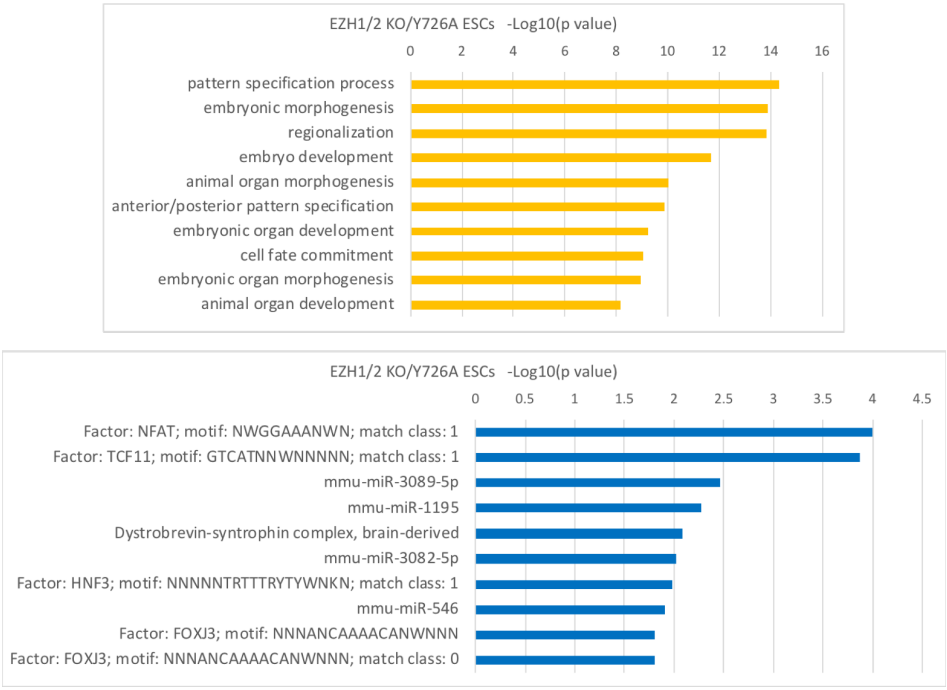


Figure 6.8 Gene ontology of differentially expressed genes in EZH1/2 KO/Y726A mESCs. GO terms enriched in genes significantly upregulated (yellow) and downregulated (blue) in EZH1/2 KO/Y726A mESCs relative to E14 control.

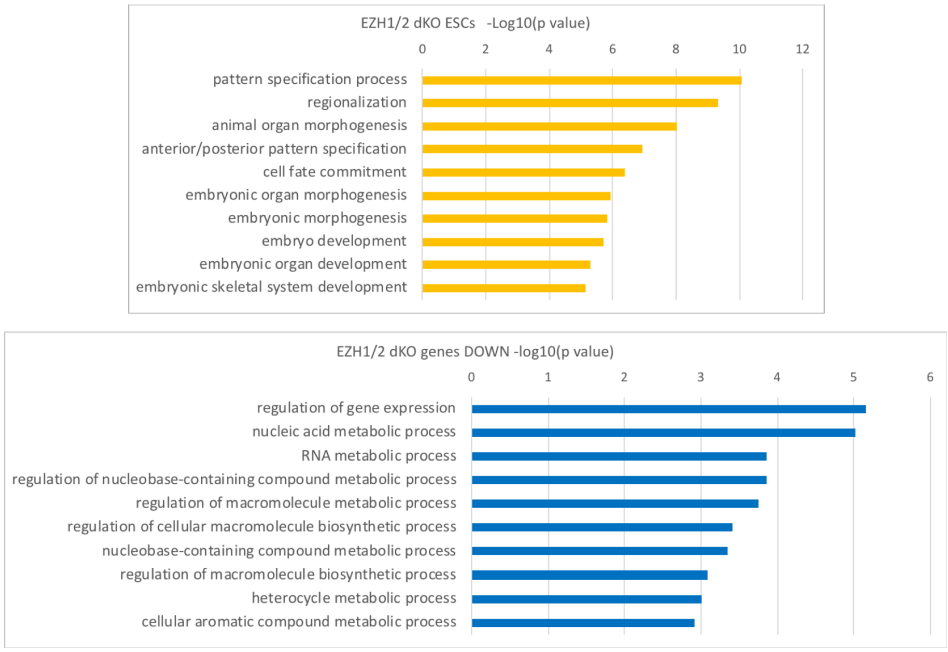


Figure 6.9 Gene ontology of differentially expressed genes in EZH1/2 dKO mESCs. GO terms enriched in genes significantly upregulated (yellow) and downregulated (blue) in EZH1/2 dKO mESCs relative to E14 control.

identified, some of which were also found in the single EZH2 Y726A background. In the EZH1/2 dKO cells, however, there was an enrichment of downregulated genes associated with gene expression and the metabolism of RNA and several other biological compounds. There were also many transcription factor binding motifs associated with the downregulated genes in EZH1/2 dKO cells, notably motifs for the binding of *YY1*, a transcription factor which had initially been associated with PRC2 recruitment (O'Carroll et al., 2001), but has since been shown to have no direct interaction with PRC2 (Sarma et al., 2008).

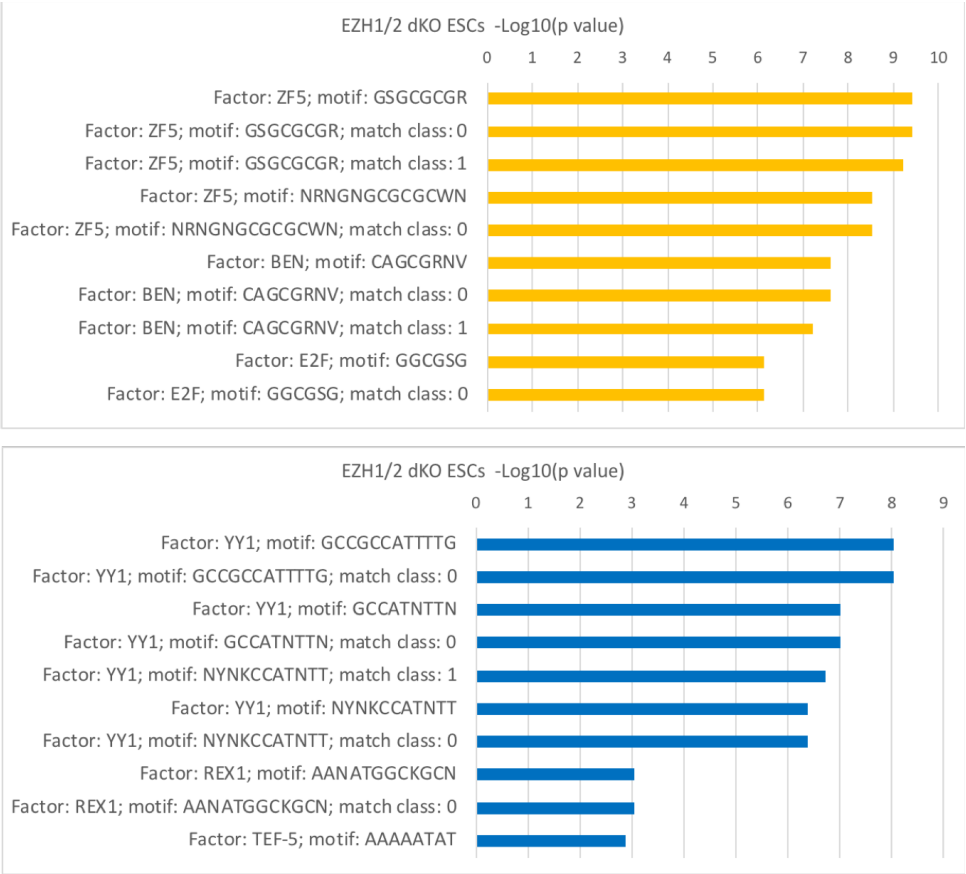


Figure 6.10 Enrichment of transcription factor binding sites in differentially expressed genes in EZH1/2 dKO mESCs. Transcription factor binding motifs enriched in genes significantly upregulated (yellow) and down regulated (blue) in EZH1/2 dKO mESCs relative to E14 control.

Although I found that a significant number of genes were uniquely differentially expressed in one or the other double mutant cell line when compared to control E14 ESCs, only a small number of genes was significantly differentially expressed when directly comparing gene expression between the two double mutant backgrounds (Fig. 6.11B). Of this group an equivalent number were upregulated and downregulated in the EZH1/2 dKO as compared to the EZH1/2

KO/Y726A, however a greater number of genes were found to be upregulated by a factor of more than 1 than were downregulated by this same factor. The GO term analysis of these genes only revealed a downregulation of genes enriched for binding motifs of the *YY1* transcription factor in the EZH1/2 dKO over the EZH1/2 KO/Y726A cells (Fig. 6.11C). The expression of *YY1* however was not found to be significantly different between the 2 double mutant cell lines, suggesting that the effects on genes which have a binding site for this factor may be down to changes in the recruitment of this factor rather than a decrease in its abundance.

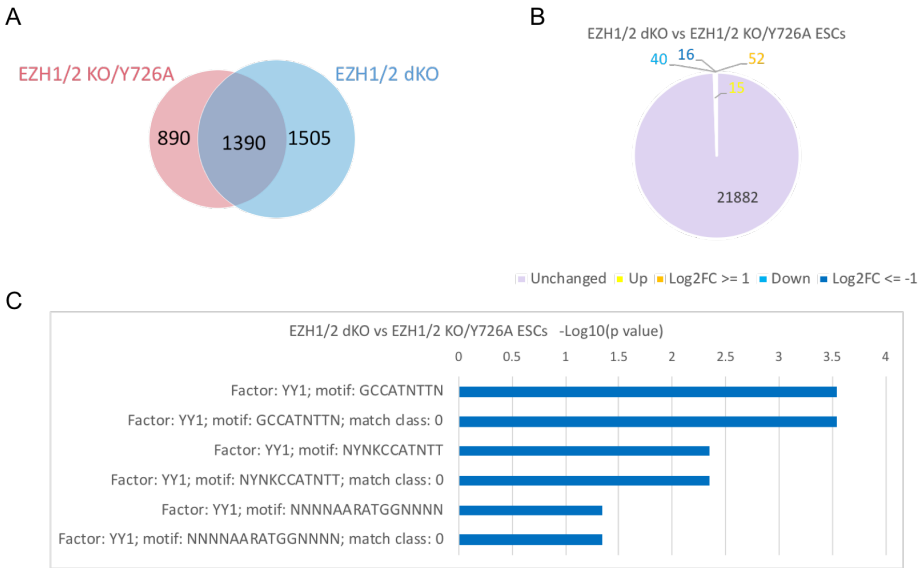


Figure 6.11 Comparison of gene expression between EZH1/2 KO/Y726A and EZH1/2 dKO mESCs. (**A.**) Venn diagram representing overlap of genes differentially expressed in EZH1/2 KO/Y726A and EZH1/2 dKO mESCs. (**B.**) Pie chart representing expression of all protein coding genes in EZH1/2 dKO relative to EZH1/2 KO/Y726A mESCs. Genes significantly upregulated in dKO cells are shown in shades of yellow, genes significantly downregulated are shown in shade of blue. Darker shades correspond to genes differentially expressed by more than a log(2) foldchange of 1 or -1 over EZH1/2 KO/Y726A cells. (**C.**) GO terms enriched in genes significantly downregulated in EZH1/2 dKO relative to EZH1/2 KO/Y726A mESCs.

I next analysed the expression profiles of selected known Polycomb targets, initially focusing on the Hox genes. Hox genes are developmental genes that play roles at several different stages of development both in *Drosophila* and in vertebrates, and have been shown to be very strongly enriched for H3K27me3 and Polycomb group proteins across broad domains in ESCs (Duboule and Dollé, 1989; Boyer et al., 2006; Zhao et al., 2007). In both EZH1/2 double

mutant backgrounds I found that the vast majority of Hox genes was significantly upregulated compared to the E14 control (Fig. 6.12). The derepression observed was very similar between EZH1/2 KO/Y726A and EZH1/2 dKO cells for all significantly differentially expressed Hox genes. In the EZH2 single mutant cells, however, none of these genes were significantly upregulated, nor was there any similar trends of gene expression to that seen in the double mutant cell lines (Fig. 6.13), suggesting that repression of Hox genes can be maintained in the absence

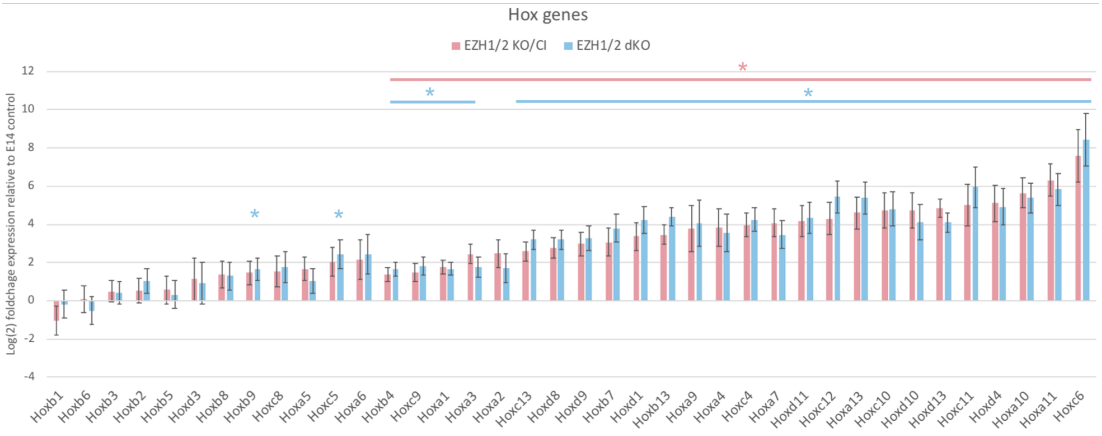


Figure 6.12 Upregulation of Hox genes in EZH1/2 double mutant mESCs. Log(2) foldchange of expression of Hox genes in EZH1/2 KO/Y726A and EZH1/2 dKO backgrounds relative to E14 control. Error bars represent standard error of Log(2) foldchange. Genes significantly differentially expressed (p value $\leq$0.05) are marked by an asterisk, pink for EZH1/2 KO/Y726A and blue for EZH1/2 dKO cells.

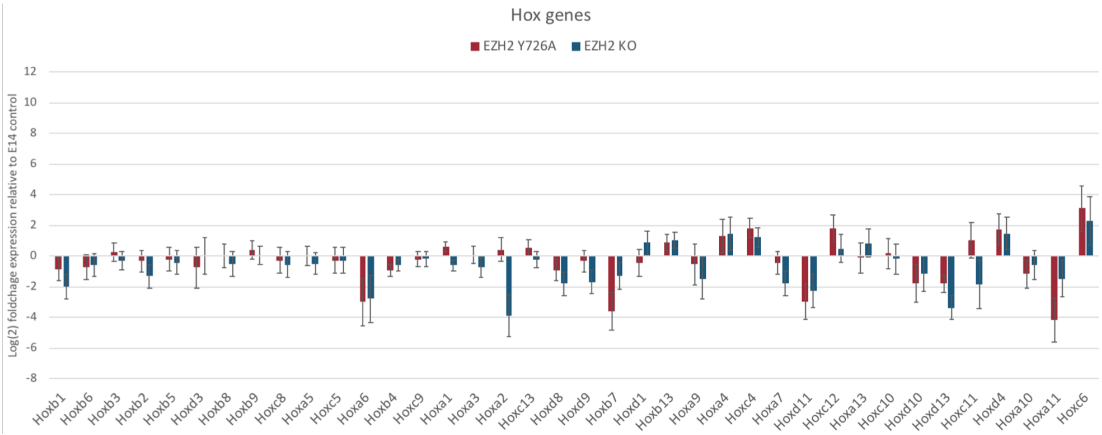


Figure 6.13 Upregulation of Hox genes in EZH2 single mutant mESCs. Log(2) foldchange of expression of Hox genes in EZH2 Y726A and EZH2 KO mESCs relative to E14 control. Error bars represent standard error of Log(2) foldchange. Of these genes none were found to be significantly differentially expressed in these cell lines.

of EZH2 but not both EZH2 and EZH1.

A similar behaviour was observed in the double mutant cells when I analysed differential expression of around 40 Polycomb target genes featuring CGIs of varying lengths in proximity to their TSS and a range of levels of H3K27me3 enrichments (based on published data), shown in Fig. 6.14. As was seen with the Hox genes, the log(2) foldchange of these genes generally correlated well between the EZH1/2 KO/Y726A and the EZH1/2 dKO cell lines.

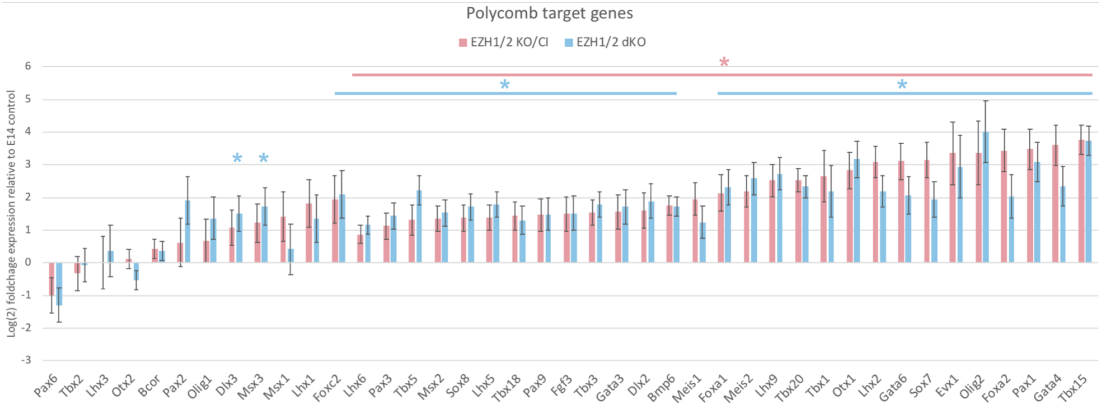


Figure 6.14 Upregulation of Polycomb target genes in EZH1/2 double mutant mESCs. Log(2) foldchange of expression of Polycomb target genes in EZH1/2 KO/Y726A and EZH1/2 dKO mESCs relative to E14 control. Error bars represent standard error of Log(2) foldchange. Genes significantly differentially expressed (p value $\leq$0.05) are marked by an asterisk, pink for EZH1/2 KO/Y726A and blue for EZH1/2 dKO cells.

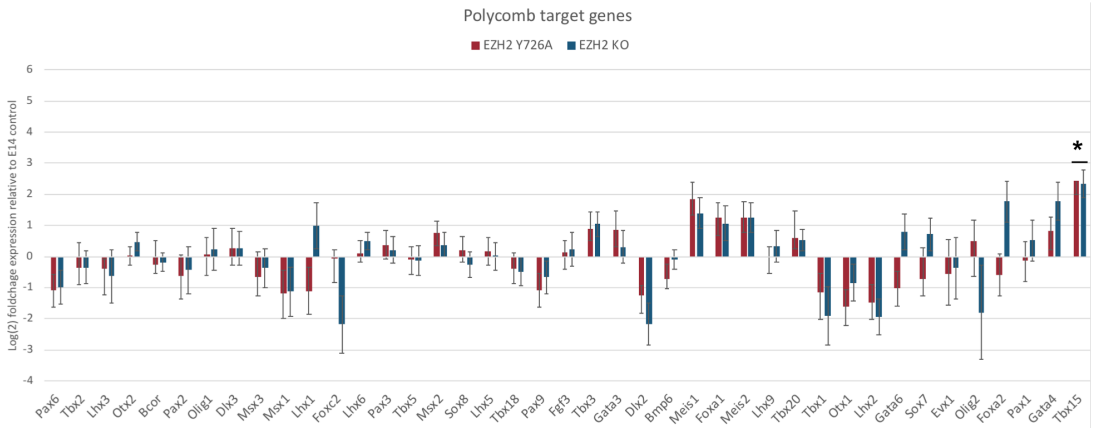


Figure 6.15 Upregulation of Polycomb target genes in EZH2 single mutant mESCs. Log(2) foldchange of expression of Polycomb target genes in EZH2 Y726A and EZH2 KO mESCs relative to E14 control. Error bars represent standard error of Log(2) foldchange. Genes significantly differentially expressed (p value $\leq$0.05) are marked by an asterisk.

These results showed that, contrary to the what was suggested by the initial RT-qPCR experiment, for the majority of overexpressed genes the two EZH1/2 double mutant cell lines induced similar levels of derepression. Again, in the EZH2 single mutants the vast majority of these genes were not found to be significantly differentially expressed (Fig. 6.15). Despite most changes in expression for the single mutants not being above the threshold of significance, I did notice that the largest changes detected in the single mutant cells were mostly seen for genes which were significantly upregulated in the double mutant cells. This trend was perhaps more evident amongst the Hox genes than the other Polycomb genes.

6.2 Conclusions

Taken together, these results demonstrated that there was a substantial difference in gene expression patterns between the EZH2 single mutant cells and the EZH1/2 double mutant cells, both with regards to the overall number of differentially expressed genes, as well as the expression of specific Polycomb targets. This further confirms, as has already been documented, that in mESCs, EZH1 is able to compensate for the majority of EZH2 functions in the regulation of gene expression. This compensation is able to occur despite the very low levels of H3K27me3 that EZH1 is able to catalyse, suggesting that high levels of this histone modification may be of minimal importance for Polycomb gene regulation in mESCs.

As expected, the most significant changes in Polycomb gene expression were observed in the EZH1/2 double mutant cell lines, with many of these genes being upregulated as the repressive PcG proteins and H3K27me3 were lost. However there were a significant number of genes that were downregulated in the EZH1/2 double mutant cell lines, which may, at first glance seem surprising. These genes may be being indirectly affected, with their downregulation due to the action of repressive factors that are direct Polycomb targets, which become expressed upon depletion of PRC2.

The experiments in this chapter also provided some information about the influence of a catalytically inactive PRC2 on gene expression. It would seem that the catalytically inactive EZH2 was still able to maintain a minor degree of gene repression as compared to the EZH2 KO, as evident from the larger number of significantly upregulated genes seen in the EZH1/2 dKO than in the

EZH1/2 KO/Y726A. Interestingly, although broadly the same genes were affected in both the KO and catalytically inactive conditions, it would seem that EZH2 Y726A was also able to prevent the silencing of certain genes that were seen to be downregulated in the EZH1/2 dKO conditions. Additionally, it appeared that the catalytically inactive EZH2 may have been able to maintain the levels of expression of genes that have a binding motif for transcription factor YY1. As compared to the effect of a total loss of H3K27me3 versus a partial loss however, the effect of maintaining a catalytically inactive EZH2 was minimal. This was illustrated by the drastic difference in the number of misregulated genes in both EZH1/2 double mutant cell lines as compared to both EZH2 single mutant cell lines, indicating that globally the catalytically inactive EZH2 was not able to contribute in a meaningful way to the silencing of its target genes.

Overall, these results point to the histone modifying activity of PRC2 as the most important factor in the maintenance of PRC2-mediated Polycomb gene silencing. It appeared that the presence of H3K27me3, even at a low level, at Polycomb target genes was sufficient to maintain the correct level of expression for the vast majority of these genes. One hypothesis to explain this is the strong decrease in RING1B enrichment at Polycomb target genes observed in both EZH1/2 double mutant but not the EZH2 single mutant backgrounds (see Chapter 4). This depletion of the catalytic subunit of PRC1 from these genes could explain their upregulation, as the this complex is thought to play an important role in gene repression by condensing chromatin and reducing the accessibility of genes to the transcription machinery (Stock et al., 2007; Zhou et al., 2008). However it has also been shown that PRC1 is not present at appreciable levels at all PRC2 target genes so it may not be the sole factor responsible for silencing of all Polycomb targets (Ku et al., 2008). Another mechanism through which H3K27me3 may be mediating repression is through inhibiting the recruitment of chromatin modifying complexes responsible for placing histone modifications associated with active genes. This could occur either through direct inhibition of the complex by the mark, or through the recruitment of additional inhibiting factors that are able to recognise H3K27me3.

Chapter 7

Differentiation of PRC2 mutant cell lines

7.1 Introduction and Aims

Although the loss of EZH2 alone is not sufficient to induce major defects in mESC self-renewal, proliferation, or gene expression patterns (Chamberlain et al., 2008; Shen et al., 2008; Riising et al., 2014) (see also previous chapter), it is known to have a considerable effect on development and ESC differentiation. The loss of EZH2 causes severe disruption of embryonic development in mice, specifically an inability to progress past gastrulation and lethality around embryonic day 8 (O'Carroll et al., 2001). A similar phenotype has also been described for knock-outs of EED and SUZ12 (Faust et al., 1998; Pasini et al., 2004). In cell culture conditions, loss of EZH2 from mouse ESCs leads to defects in differentiation. Typically these defects are characterised by a failure to repress pluripotency factors, aberrant expression of genes involved in lineage specification, and failure to upregulate genes involved in key pathways required for cell fate commitment (Shen et al., 2008). As with embryo development, these differentiation defects have also been observed in the absence of other PRC2 subunits (Chamberlain et al., 2008; Pasini et al., 2007). In human cells the loss of EZH2 leads to a more severe phenotype, as ESCs exhibit defects in differentiation similar to those seen in mice, although in this case there is no failure to repress pluripotency factors, but also have a reduced capacity for self-renewal (Collinson et al., 2016; Shan et al., 2017).

Although in undifferentiated ESCs I found there to be no discernible difference between the EZH2 KO and EZH2 Y726A cell lines, with the following experiments I aimed to ascertain whether the catalytically inactive EZH2 main have a significant effect on gene expression throughout differentiation as compared to the knockout. I also sought to gain some insight into which stages of differentiation PRC2 may be the most vital for and whether its full catalytic activity is required for all of its function in gene regulation during this process. I hypothesised that although the catalytically inactive form of EZH2 was unlikely to restore full differentiation capacity to the cells, it may allow the progression of differentiation to a more mature state than that seen in the EZH2 KO cells.

The differentiation protocol used in this chapter is based on the protocol described by Bibel et al. (Bibel et al., 2007) that generates a uniform population of glutamatergic neurons. This protocol has been used in studies of Polycomb proteins (Mohn et al., 2008; Schmitz et al., 2011) and is a straightforward, well characterised procedure which allowed me to examine the requirements of PRC2 throughout a multi-step differentiation process. As this protocol involves a progression through several states of differentiation it allowed me to assess the contribution of PRC2 not only to the early stages of differentiation during the formation of embryoid bodies (EBs), but also to the later stages in the transition of neural precursors to the terminally differentiated neurons. The expression of several marker genes throughout the protocol has been described in detail, allowing me to verify the differentiation process was working correctly in my hands. These markers were also of use to begin to pinpoint the stages at which the EZH2 deficiency becomes antagonistic to the correct regulation of gene expression. It should be noted however, that although the expression patterns of some specific markers are well characterised, there remains much to be learned about the regulation of cell fate decisions from analysing global gene expression throughout differentiation.

The immunofluorescence staining of cells in this chapter was performed with the help of Katy McLaughlin. The quality control and processing of RNA-seq raw data was performed under my direction by Shaun Webb.

7.2 Overview of neuronal differentiation protocol

I performed neuronal differentiation assays with both the E14 control cells and the PRC2 mutant cells to test whether presence alone or catalytic activity of EZH2 was required for this process. I followed the progression of differentiation first by monitoring the morphology of cells as they went from ESCs to EB, neural progenitor, and finally terminally differentiated neurons, before analysing gene expression changes by RT-qPCR and RNA-seq. An outline of the differentiation protocol along with expected expression behaviour of marker genes is shown in Fig. 7.1. To determine how singular loss of EZH2 compares to loss of both EZH1 and EZH2, I performed differentiation experiments with both single and double mutant cell lines.

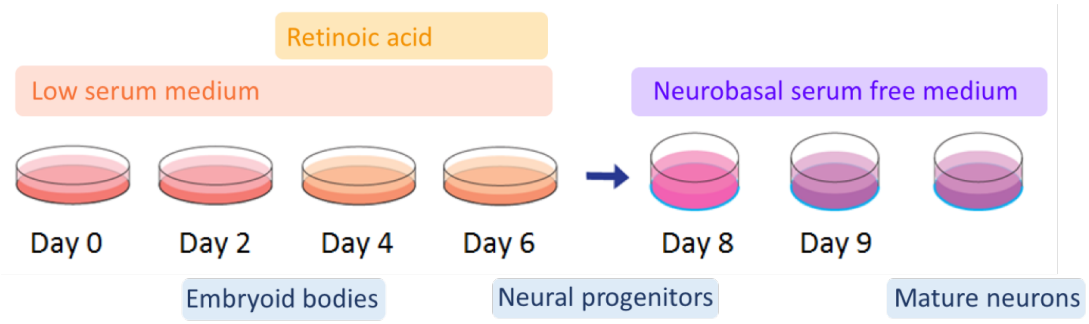


Figure 7.1 Scheme of the neural differentiation protocol. Detail of media changes and cell types at each stage of the differentiation.

7.3 Cell morphology of PRC2 mutant cell lines throughout differentiation

Throughout the first stages of differentiation the PRC2 mutant cell lines behaved similarly to the E14 control cells. The EBs were of a similar size and shape and appeared to grow in size at a similar rate (data not shown). After the addition of retinoic acid (RA), however, obvious differences in morphology became apparent, between the E14 control and EZH1/2 double mutant cell lines in particular. As can be seen in Fig. 7.2 where the E14 control line started to produce more smooth and evenly shaped EBs by day 6, the EBs in the double mutant lines were smaller,

more unevenly shaped with less defined outlines and even occasionally appeared to be hollow. This phenotype was not as pronounced in the EZH2 single mutant cell lines, for which the EBs formed were slightly smaller than the E14 control and had a slightly less smooth edge (Fig. 7.2).

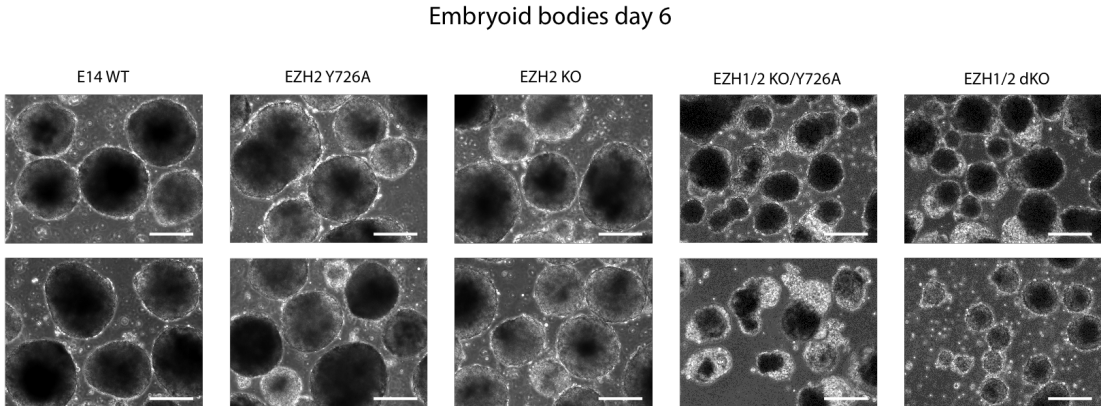


Figure 7.2 Embryoid bodies on day 6 of differentiation. Brightfield imaging at 10x magnification of embryoid bodies on day 6 of differentiation. Two representative images (upper an lower panels) of EB morphology of E14 control (E14 WT) and EZH2 single and double mutant cell lines across at least 3 separate differentiation experiments. Scale bar corresponds to 100 µm.

After trypsinisation on day 8, cells were counted and the same number in each line plated onto PDL-laminin coated plates (see Fig. 7.1). A few hours after plating cell morphology was similar for the EZH2 single mutants cell lines and the E14 control. However, in the double mutants there already appeared to be a greater level of cell death, and the remaining live cells partially failed to properly adhere to the plate. By day 9 the differences between E14 control and EZH2 single mutant cell lines started to become evident (Fig. 7.3). Whereas the E14 control cells were beginning to form the typical neurite structure seen in neuronal cells, this was not apparent in either EZH2 single mutant background (Fig. 7.3), even though a similar number of viable cells were present on the plate in the EZH2 single mutants as compared to the E14 control. In contrast, this was not the case for the EZH1/2 double mutant cell lines, which not only failed to undergo the morphological changes seen for E14 control cells, but were also subject to significant rates of cell death by day 9. These low cell numbers meant that I was unable to obtain RNA samples for the double mutant cell lines past day 8 of the differentiation protocol.

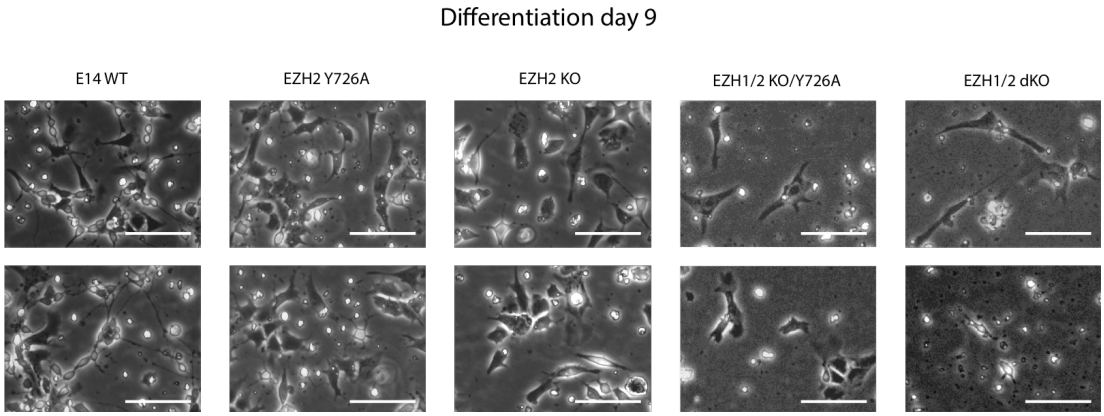


Figure 7.3 Cell morphology on day 9 of differentiation. Brightfield imaging at 10x magnification of E14 control (E14 WT) and EZH2 single and double mutant cell lines on day 9 of differentiation. Cells plated on PDL-laminin coated plates. Two representative images (upper an lower panels) of cell morphology as seen across at least 3 separate differentiation experiments. Scale bar corresponds to 50 µm.

Throughout the next week of differentiation, the E14 control cells maintained and further developed the network of neuronal cells, whereas both EZH2 single mutants failed to form proper neuronal protrusions and networks. There were, however, no obvious signs of significant cell death in these lines. Although they did not appear neuronal from a morphological point of view, they were able to grow and even proliferate in the neurobasal B27 supplemented media as evidenced by the increase in cell density on the plates over time for these cell lines. Based on their morphology, the EZH2 single mutant cells at this stage appeared to be a mixed population of cell types (Fig. 7.4).

This morphological assessment of neuronal differentiation in each cell line revealed that none of the PRC2 mutant cell lines were able to differentiate in the same manner as the E14 control. Although the EZH2 single mutants were able to survive throughout the entire course of differentiation, they never took on a morphology resembling the network of neurites typically formed by neuronal cells, as was seen in the E14 control cells. The EZH1/2 double mutants were even less capable to carry out this neuronal differentiation, with very few of these cells even surviving through to the later stages.

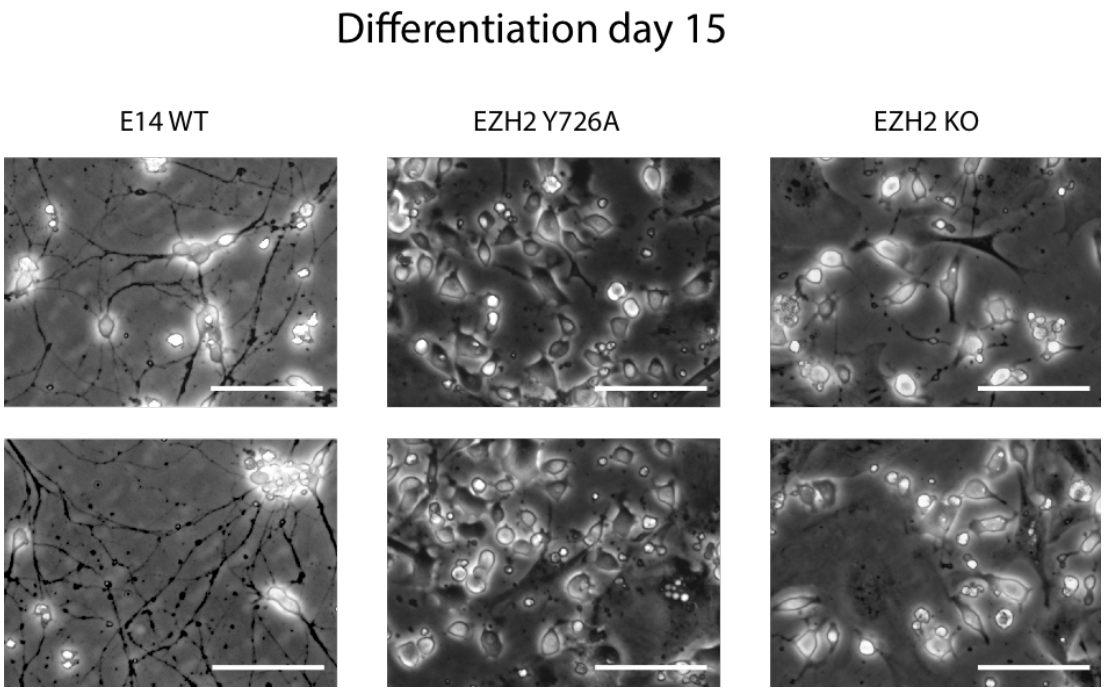


Figure 7.4 Cell morphology on day 15 of differentiation. Brightfield imaging at 10x magnification of E14 control (E14 WT) and EZH2 single and double mutant cell lines on day 15 of differentiation. Cells plated on PDL-laminin coated plates. Two representative images (upper an lower panels) of cell morphology as seen across at least 3 separate differentiation experiments. Scale bar corresponds to 50 µm.

7.4 Analysis of gene expression in PRC2 single mutant cell lines throughout differentiation

7.4.1 RT-qPCR analysis of differentiation markers

RT-qPCR was performed at different time points throughout differentiation for select marker genes for different stages of differentiation, including *Pax6* and *Fapb7* which are two marker genes used to identify cells undergoing correct neural differentiation following the protocol described by Bibel et al. 2007 (Bibel et al., 2007).

As expected, *Otx2*, a TF induced early in EB formation was found to be upregulated by day 2 in the E14 control cells and was later repressed by day 6, likely in response to the RA treatment (Fig. 7.5). This gene was upregulated and silenced in a timely manner and to levels close to those observed in the E14

control cells for both EZH2 single mutant cell lines, suggesting that early stages of differentiation could be unaffected by both single mutations. The *Pax6* gene, a TF that serves as a marker of neural precursors, was found to be expressed from day 6 onwards in the E14 control cells. Despite a slight upregulation of this gene in the EZH2 single mutants that followed the temporal pattern of the E14 control cells, expression failed to reach the levels seen in the E14 control cells (Fig. 7.6), suggesting increasing defects in these cell lines as differentiation progresses. The *Fabp7* gene is a fatty acid binding-protein essential for early brain development which is induced upon expression of *Pax6* and marks more terminally differentiated neurons (Arai, 2005). In the E14 control cells *Fabp7* was upregulated from day 8 of differentiation onwards, reaching very high levels of induction by day 15. However it did not ever become expressed in either of the EZH2 single mutants.

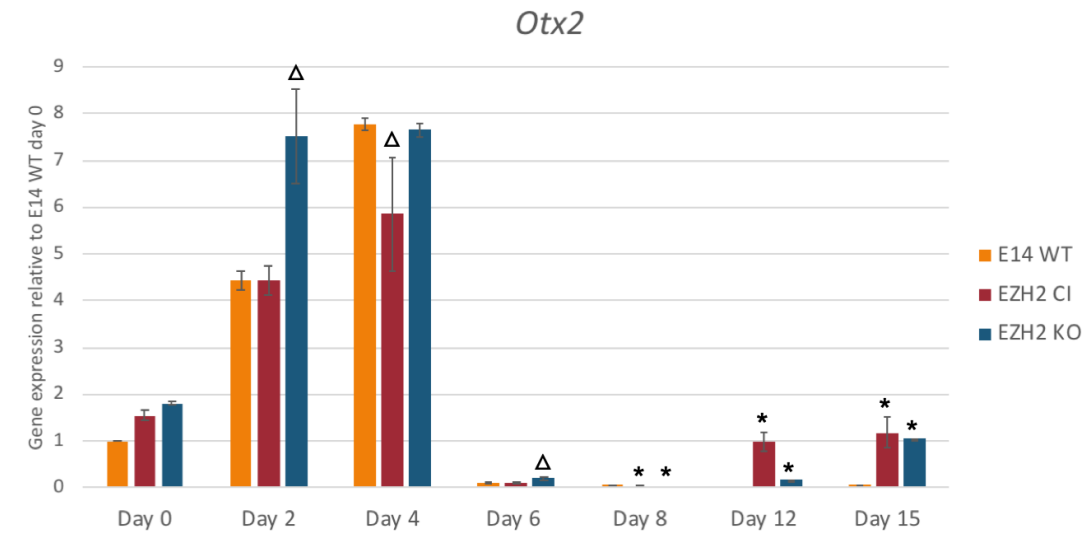


Figure 7.5 Expression of *Otx2* throughout differentiation. Bar chart showing expression of early differentiation marker *Otx2* in E14 control (E14 WT) , EZH2 Y726A (EZH2 CI) and EZH2 KO cell lines throughout differentiation. Data was generated by RT-qPCR and normalised to expression of *Gapdh*, expression was then expressed as fold change relative to E14 control on day 0. Error bars represent SEM across 3 biological replicates. * denotes a significant difference to WT with a p value of less than 0.01, and Δ a p value of less than 0.05.

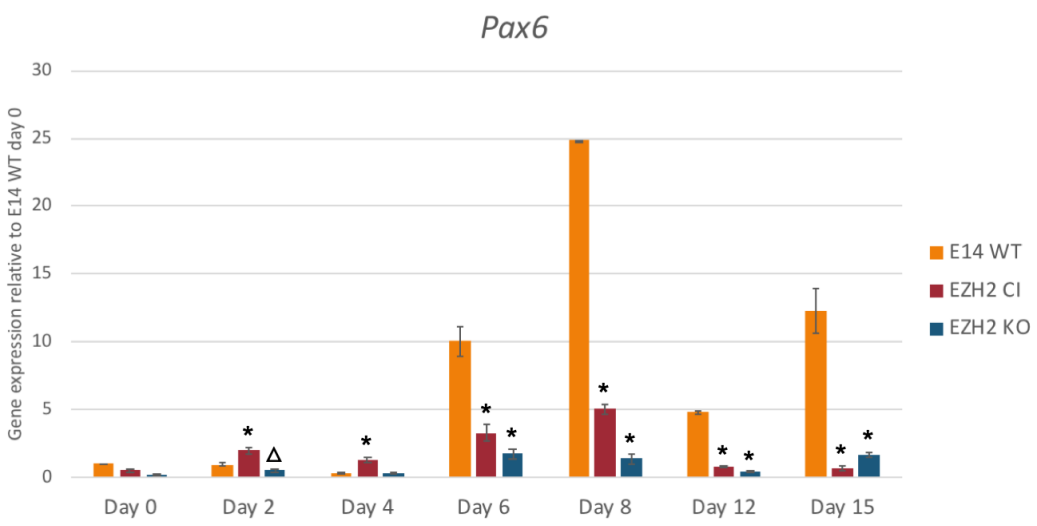


Figure 7.6 Expression of *Pax6* throughout differentiation. Bar chart showing expression of neural-progenitor marker *Pax6* in E14 control (E14 WT), EZH2 Y726A (EZH2 CI) and EZH2 KO throughout differentiation. Data was generated by RT-qPCR and normalised to expression of *Gapdh*, expression was then expressed as fold change relative to E14 control on day 0. Error bars represent SEM across 3 biological replicates. * denotes a significant difference to WT with a p value of less than 0.01, and Δ a p value of less than 0.05.

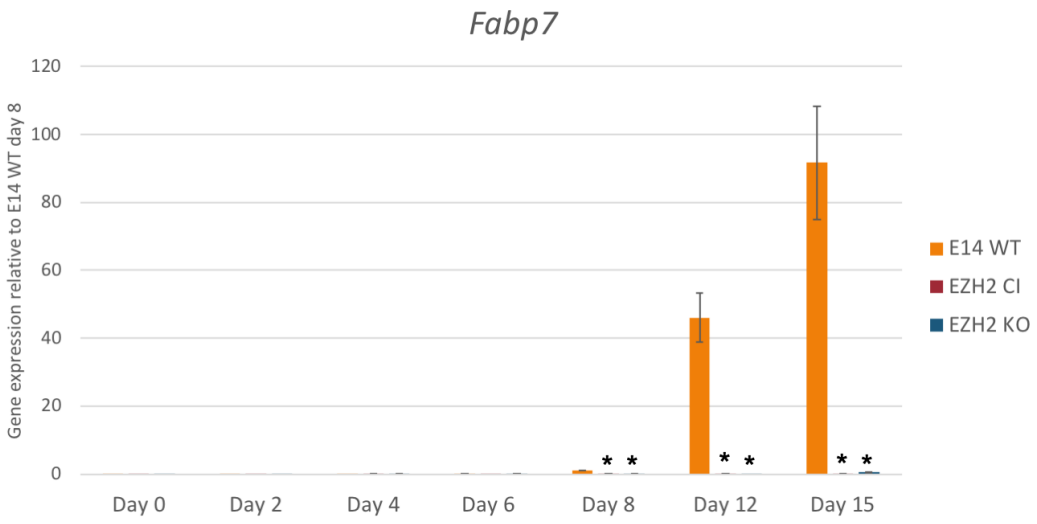


Figure 7.7 Expression of *Fabp7* throughout differentiation. Bar chart showing expression of neuronal marker *Fabp7* in E14 control (E14 WT), EZH2 Y726A (EZH2 CI) and EZH2 KO throughout differentiation. Data was generated by RT-qPCR and normalised to expression of *Gapdh*, expression was then expressed as fold change relative to E14 WT on day 8. Error bars represent SEM across 3 biological replicates. * denotes a significant difference to WT with a p value of less than 0.01

As can be seen by immunofluorescence staining in Fig. 7.8, the E14 control cells expressed high levels of neuronal marker *Tuj1* throughout their cell body by day 15. As was shown by the brightfield images of cells on day 15, the EZH2 single mutants did not form the neurite network characteristic of neuronal cells (Fig. 7.4). Immunofluorescence staining of cells on day 15 showed that, in addition to not displaying a neuronal morphology, the EZH2 single mutants also failed to express *Tuj1*. Furthermore, there appeared be a higher degree of *Oct4* expression in these cells than in the E14 control (Fig. 7.8).

These results indicated that although genes such as *Otx2* were able to be up- and subsequently downregulated in the earlier stages of differentiation in both EZH2 single mutant backgrounds, neither of them were able to initiate the expression of neural lineage programmes at the EB stage, eventually leading to their failure to produce mature neurons.

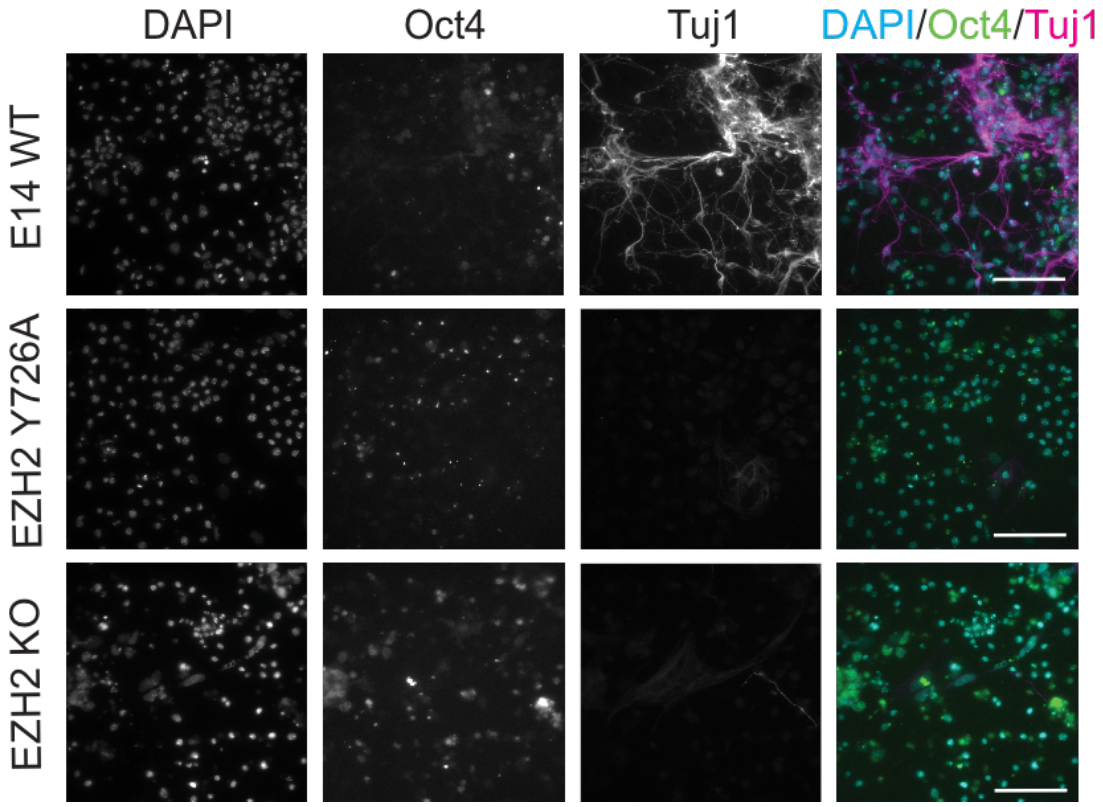


Figure 7.8 Immunofluorescence staining of OCT4 and TUJ1 on day 15 of differentiation. DAPI (far left), OCT4 (left), and TUJ1(right) staining of E14 control, EZH2 Y726A and EZH2 KO cells on day 15 of differentiation. Merge is shown in panels on far right, DAPI in cyan, OCT4 in green and TUJ1 in magenta. Scale bar corresponds to 100 µm

7.5 Analysis of genome-wide gene expression patterns during differentiation by RNA-seq

To further explore the gene expression dynamics throughout differentiation, RNA-seq was performed at 3 time points throughout differentiation. Samples were taken on day 2, day 8 and day 15 for E14 control, EZH2 Y726A, EZH2 KO and EZH1/2 dKO cells using 3 separate clones of each genotype as replicates for each time point. These time points were chosen with the aim to capture gene expression profiles and dynamics across the full differentiation protocol in order to encompass both early and late stage events. By comparing the EZH2 single mutant cell lines to each other I aimed to gain insight into any potential catalysis-independent functions of EZH2 throughout differentiation. By including the EZH1/2 dKO cells in this assay as a comparison to the EZH2 single mutant cells I also aimed to shed some light on the functions that EZH1 is able to carry out throughout differentiation in the absence of EZH2.

7.5.1 Gene expression dynamics throughout differentiation

I will first discuss the dynamics in gene expression that I observed in the E14 control cell line to establish what was expected at each stage of differentiation in terms of pathways and families of genes that are up- and downregulated.

Day 2 EB stage

Although the samples for day 0 and day 2 of differentiation cluster together very closely in principle component analysis (PCA) compared to the later days of differentiation (see Fig. 7.20), a considerable number of genes nonetheless changed their expression from day 0 to day 2. Approximately 12% of protein coding genes were differentially expressed by day 2, with around 1100 genes upregulated and 1570 downregulated. The analysis of GO terms of upregulated genes identified an enrichment of genes involved in cell communication and signalling as well as the metabolism of sterols, cholesterol and alcohol. Genes downregulated at this time were found to be mostly involved in the regulation of gene expression, as well as RNA metabolism and processing (Fig. 7.9).

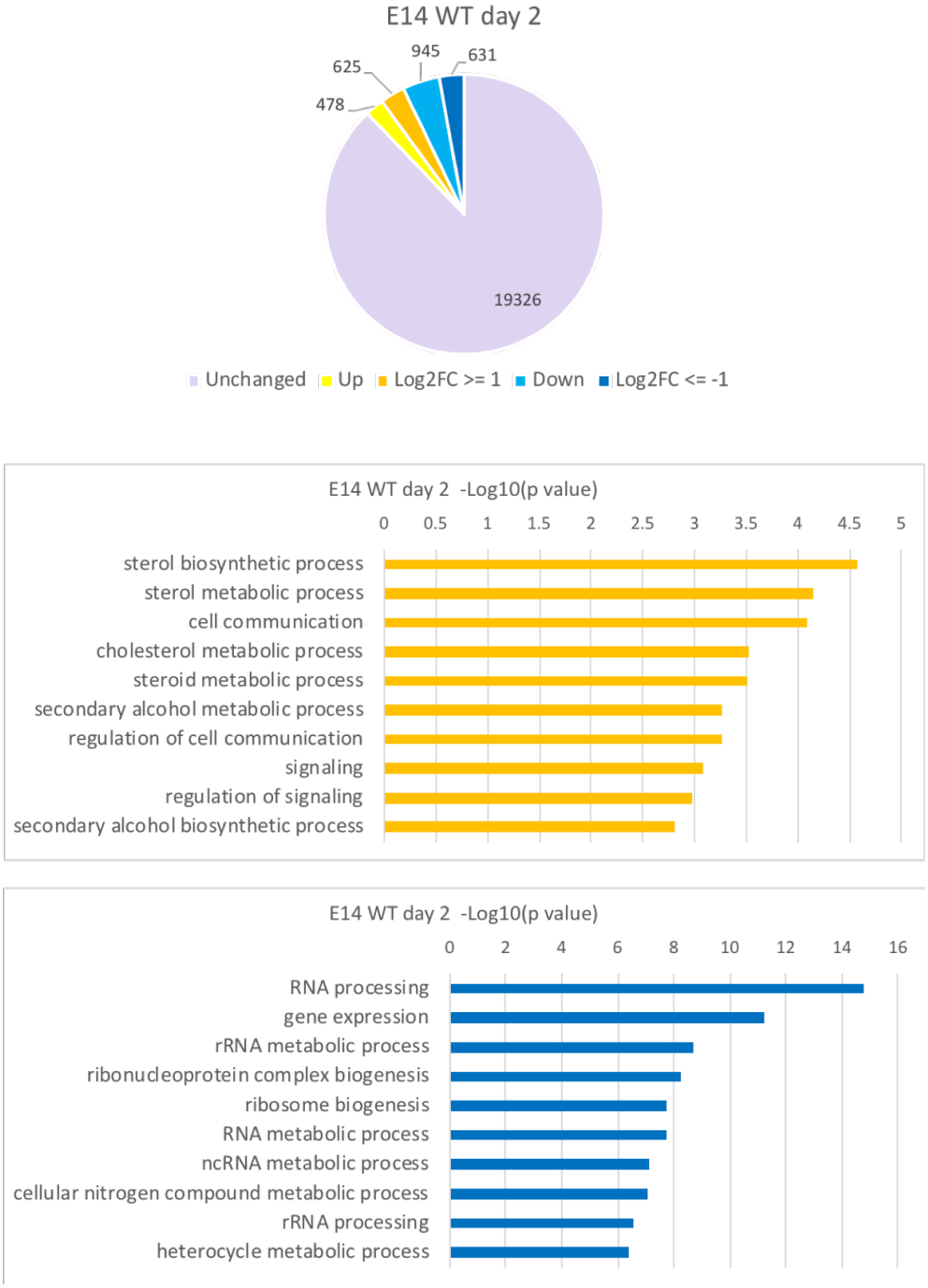


Figure 7.9 Differential gene expression in day 2 relative to undifferentiated mESCs for E14 control. Upper panel - Pie chart representing differential expression of all protein coding genes in day 2 E14 control (E14 WT) cells relative to undifferentiated mESCs. Genes significantly upregulated are shown in shades of yellow, genes significantly downregulated are shown in shades of blue. Darker shades correspond to genes differentially expressed by more than a log(2) foldchange of 1 or -1 over E14 control. Lower panels - GO term enrichment of significantly upregulated (yellow) and downregulated (blue) genes.

When analysing the association of these genes with specific TF families in order to begin understanding how these genes might be regulated, I observed that genes upregulated on day 2 were enriched for TF binding motifs that were GC rich, many of which corresponded to the binding motifs for factors from the Sp family. Amongst genes that were downregulated there was an enrichment of binding motifs of several different transcription factors. Of these binding motifs many of them had a CCGGAA motif in common (Fig. 7.10).

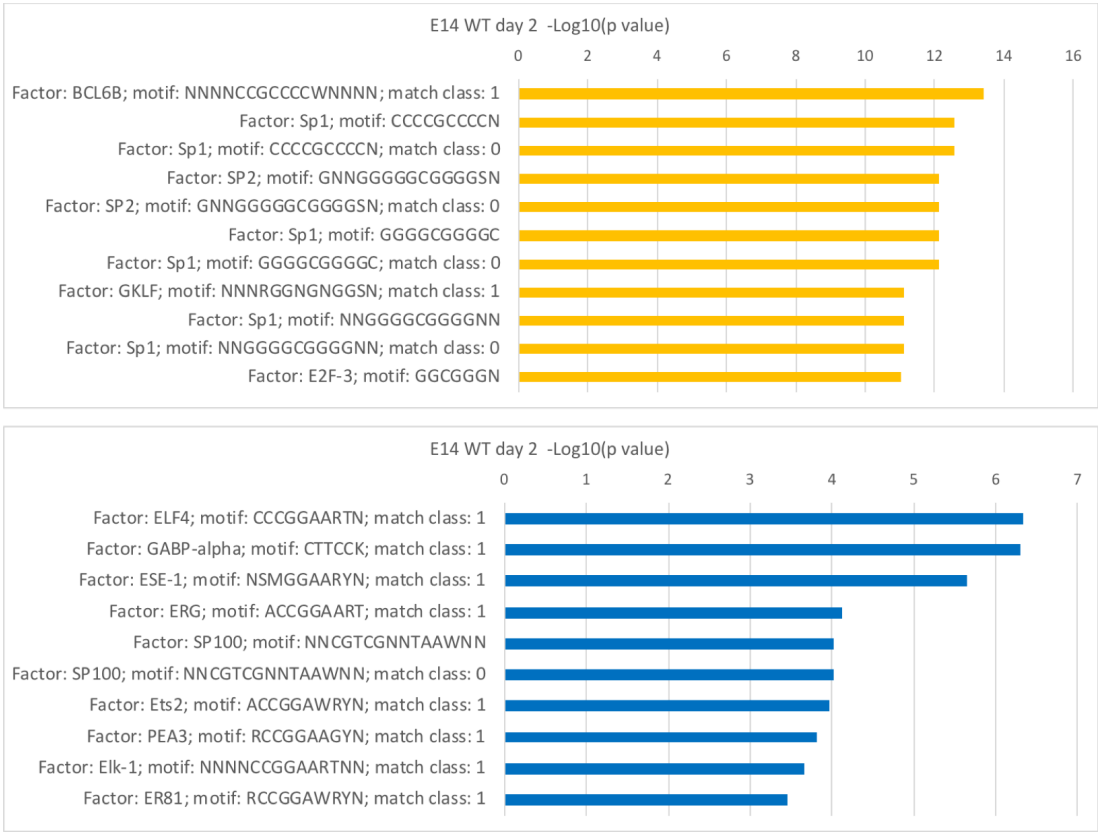


Figure 7.10 Enrichment of TF motifs in differentially expressed genes in day 2 E14 control. TF binding motifs enriched in genes significantly upregulated (yellow) and down regulated (blue) in day 2 E14 control cells (E14 WT) relative to undifferentiated mESCs.

I next assessed the expression of a selection of specific marker genes at this time point. As expected from the findings of previous studies, the withdrawal of LIF at this first stage of differentiation led to the downregulation of genes such as *Klf4*, *Stat3* and *Tbx3* (Fig. 7.11), all genes involved in the maintenance of self-renewal and an undifferentiated state in ESCs (Trouillas et al., 2009; Zhang et al., 2010; Wang et al., 2017; Lu et al., 2011). This also led to an upregulation of genes such as *Otx2* and *Fgf5* (Fig. 7.11), both markers of early ectoderm (Haub and Goldfarb, 1991; Gammill and Sive, 2001).

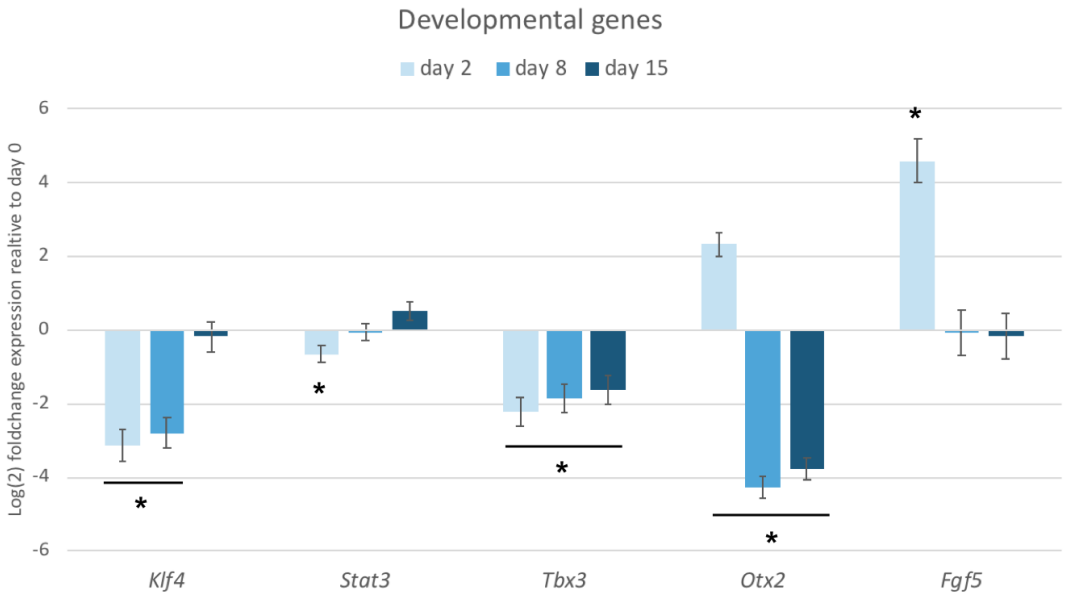


Figure 7.11 Expression of developmental genes in E14 control throughout differentiation. Log(2) foldchange of expression of genes known to be affected by withdrawal of LIF in E14 control cells at each timepoint relative to undifferentiated ESCs. Significant changes in expression (p value ≤ 0.05) are marked by an asterisk.

Pluripotency factors *Oct4*, *Nanog*, and *Sox2* were found to still be expressed at this stage of differentiation, albeit at reduced levels (Fig. 7.14) This was to be expected as their expression is known to be at least partially maintained in the first few days of EB formation (Trouillas et al., 2009).

Day 8 Transition to neural progenitor state

By day 8 approximately 40% of protein coding genes were differentially expressed as compared to day 2, with a similar number of genes upregulated as downregulated. The gene expression pattern observed on day 8 appears to reflect the treatment of cells with retinoic acid and their transition into the neural progenitor state. As expected, amongst upregulated genes there was a strong enrichment of genes involved in nervous system development and cell differentiation (Fig. 7.12). There was a strong downregulation of genes involved in translation and peptide biosynthesis, RNA processing and gene expression as well as several metabolism pathways such as that of nitrogen compounds and RNA, indicating a switch from a highly proliferative to a more neuronal phenotype (Fig. 7.12).

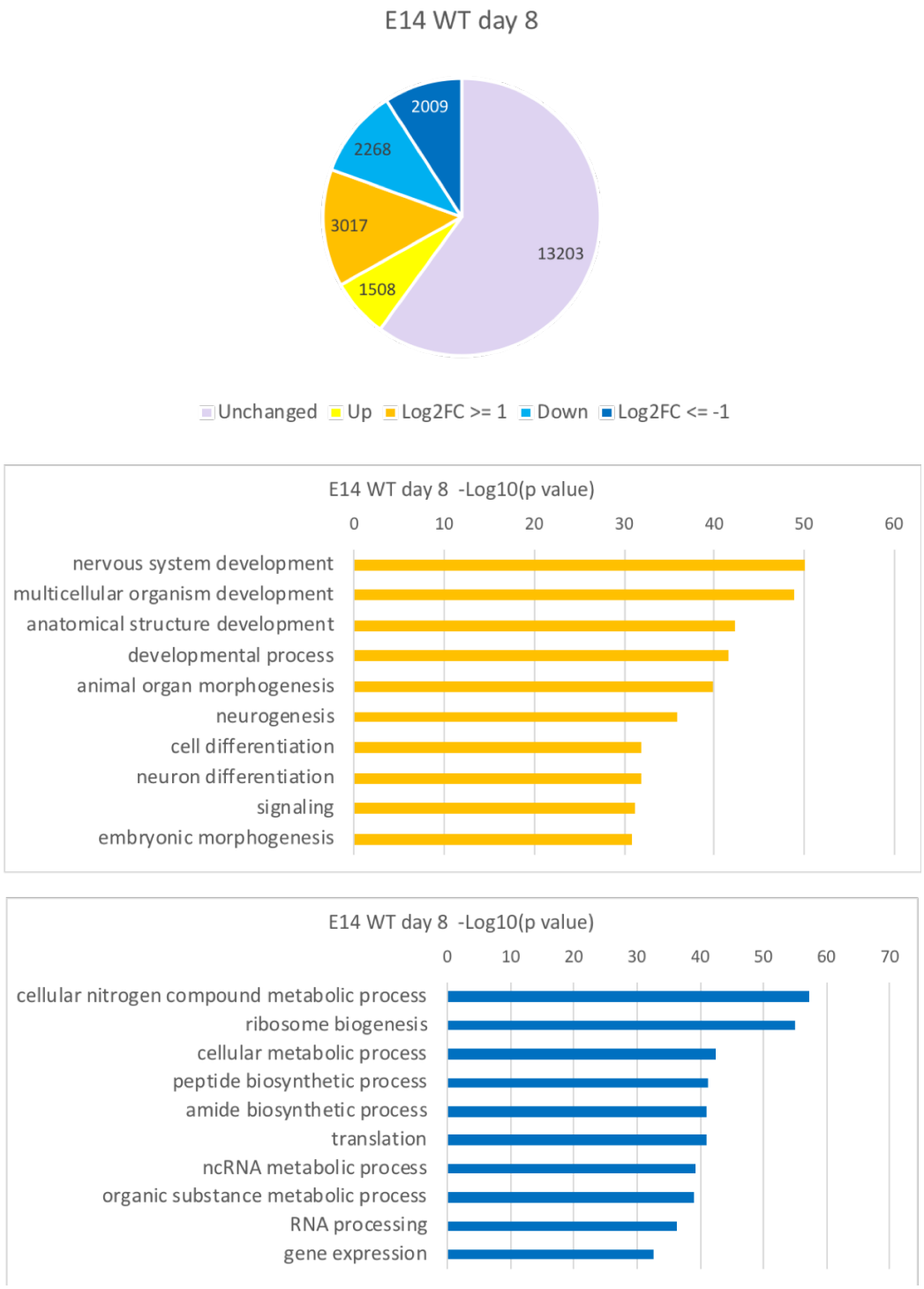


Figure 7.12 Differential gene expression in day 8 E14 control cells relative to day 2. Top panel - Pie chart representing expression of all protein coding genes in day 8 E14 WT cells relative to day 2. Genes significantly upregulated are shown in shades of yellow, genes significantly downregulated are shown in shades of blue. Darker shades correspond to genes differentially expressed by more than a log(2) foldchange of 1 or -1 over day 2 E14 control. Lower panels - GO term enrichment of significantly upregulated (yellow) and downregulated (blue) genes.

As was seen on day 2, there was an enrichment of GC rich TF motifs amongst the upregulated genes (Fig. 7.13). These binding motifs were however associated with a wider variety of TFs than seen on day 2. Of the transcription factor binding motifs enriched in the downregulated genes, many of those most enriched contained a CACGTG motif. Of the upregulated genes there is an enrichment of the binding motif for GKLF. This is another name for the *Klf4* gene, and though its expression decreases in early stage embryoid bodies, it is re-expressed in neural progenitor cells (Qin and Zhang, 2012). Also enriched amongst upregulated genes were binding motifs for Kaiso, RNF96, AP2 and ZF5. Kaiso is a methyl-CpG binding protein that has been suggested to play a role in neural development (Martín Caballero et al., 2009). RNF96, otherwise known as Trim28 is a transcriptional repressor, involved in the maintenance of stem cell pluripotency (Fazzio et al., 2008). The AP2 transcription factors are expressed in primitive ectoderm and are important for the development of the neural crest (Eckert et al., 2005). ZF5, also known as Zbtb14, has been shown to act both as a transcriptional repressor and activator and promotes the development of neural tissues (Takebayashi-Suzuki et al., 2018). Amongst the TFs for which there was an enrichment of the corresponding binding factors in the downregulated genes c-Myc was identified, one of the Yamanaka reprogramming factors and a major regulator of stem cell identity (Takahashi and Yamanaka, 2006). These enrichments taken together again demonstrated a transition of cell state from pluripotent stem cell to neural lineage at this stage of differentiation.

As expected at this time in the differentiation process there was a strong induction of transcription factors and markers of neural progenitors such as *Pax6*, shown in Fig. 7.15. This was accompanied by the upregulation of additional factors such as *Glast*, *Fabp7* and *Nestin*, all of which are associated with radial glial progenitor cells which *in vivo*, give rise to a large portion of neurons in the central nervous system (CNS) (Bibel et al., 2007; Barry et al., 2014). It also became evident that pluripotency factors were being repressed markedly more strongly compared to day 2, as *Oct4*, *Nanog* and *Sox2* were all significantly downregulated at this stage of differentiation (Fig. 7.14). Additionally, amongst genes downregulated were many of those that had been induced upon withdrawal of LIF such as *Fgf5* and *Otx2* (Fig. 7.11). The treatment of cells with retinoic acid also resulted in the upregulation of genes involved in RA signalling, such as *Rarb* and *Rxra*, nuclear receptors of RA, and *Crabp1/2*, RA binding proteins involved in its transport into cells (Fig. 7.15).

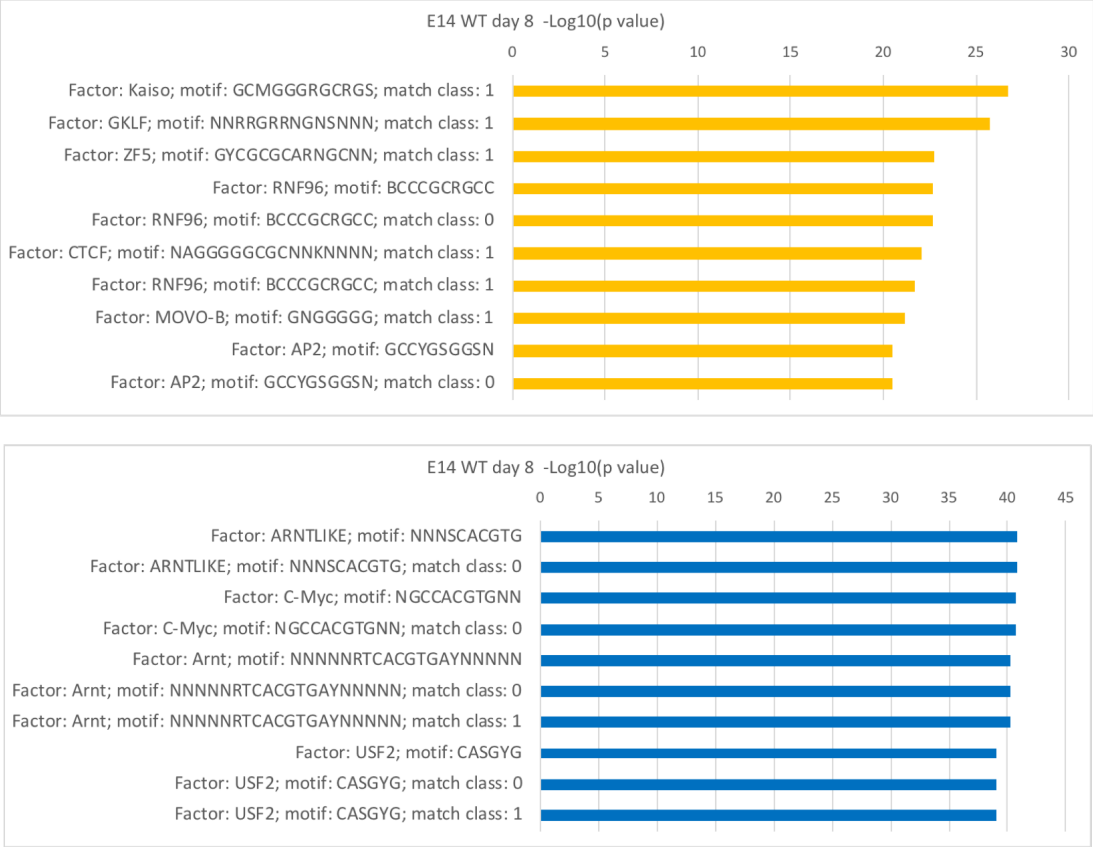


Figure 7.13 Enrichment of transcription factor binding sites in differentially expressed genes in day 8 E14 control cells. Transcription factor binding motifs enriched in genes significantly upregulated (yellow, top panel) and down regulated (blue, bottom panel) in day 8 E14 control relative to day 2.

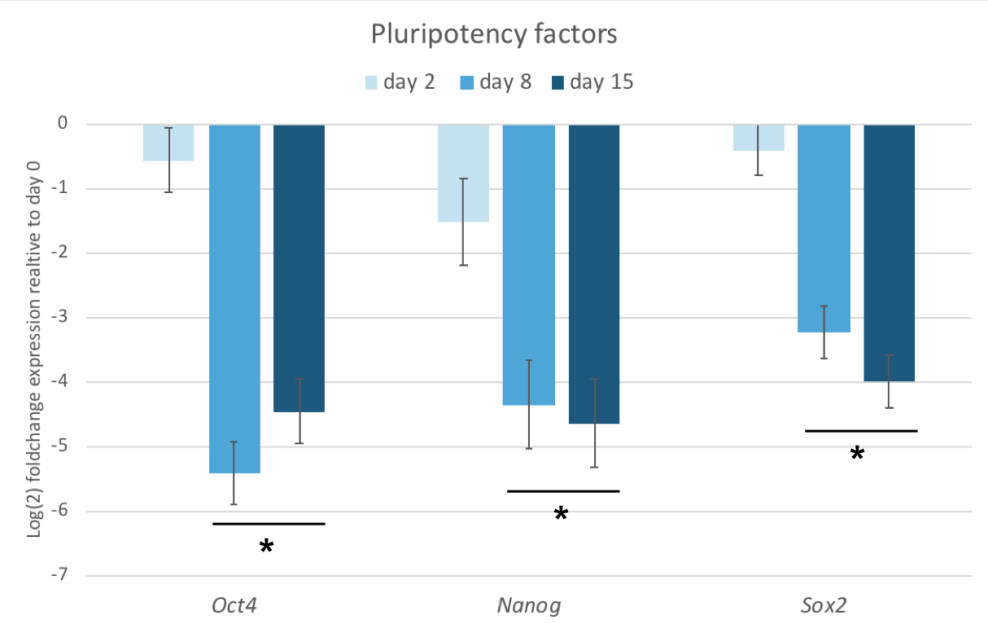


Figure 7.14 Expression of key pluripotency factors in E14 control throughout differentiation. Log(2) foldchange of expression of pluripotency factors in E14 control cells at each timepoint relative to undifferentiated ESCs. Significant changes in expression (p value $\leq$0.05) are marked by an asterisk.

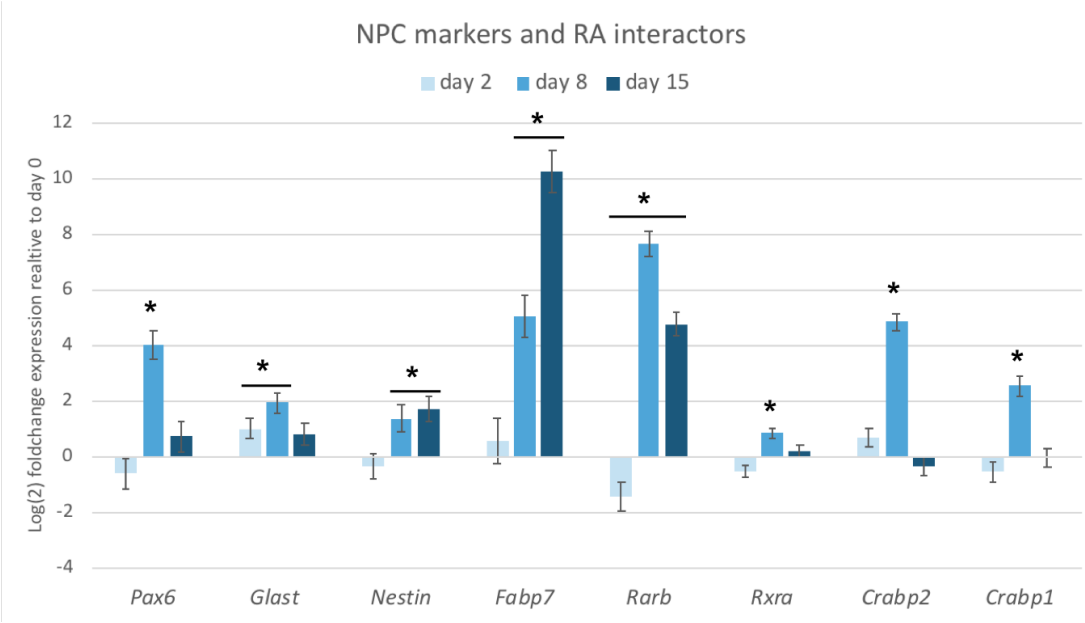


Figure 7.15 Expression of NPC markers and retinoic acid interactors in E14 control cells throughout differentiation. Log(2) foldchange of expression of neural progenitor marker genes and retinoic acid interacting genes in E14 control (E14 WT) cells at each timepoint relative to undifferentiated E14 control. Significant changes in expression (p value $\leq$0.05) are marked by an asterisk.

Day 15 Terminally differentiated neuron stage

Considerable changes in gene expression were observed between the neural progenitors of day 8 to the adherent, neurite-displaying cells of day 15. Nearly 4310, or close to 20% of all protein coding genes were significantly upregulated while 3200 or almost 15% were downregulated relative to the day 8 E14 control cells (Fig. 7.16). As cells developed into more mature neurons their proliferative capacity was reduced, which was reflected in the GO term enrichments of downregulated genes being associated with cell cycle and mitosis. Chromatin organisation as well as DNA processing and packaging were also enriched among downregulated genes, as would be expected in non-cycling cells that no longer need to condense chromosomes in preparation for cell division. The increasingly differentiated state of cells at this stage was reflected in the continued upregulation on day 15 of genes associated with neuronal development as well as synapse signalling and the formation of neuronal projections.

For upregulated genes, TF motifs were again found to be enriched for CG rich sequences. This was also true, although to a lesser extent, in genes downregulated for which the transcription factors with the most enriched binding motifs all belonged to the E2F family. The E2F family of transcription factors was also present amongst those with binding motifs most enriched in upregulated genes (Fig. 7.17). This family of transcription factors is a major actor in the control of cell proliferation and has been found to act both as an activator and a repressor of transcription (Müller and Helin, 2000). Its presence at both up- and downregulated genes again suggests changes in the rate of cell cycling.

Reflecting the state of the neurons at this stage of differentiation, there was an increase in the expression of genes such as *Fabp7* and *TrkB* as well as *Vgat* and *Vglut*, transporters of GABA and glutamate, respectively, (Fig. 7.18), identifying these cells as GABAergic and glutamatergic neurons (Bibel et al., 2004). *Pax 6* was downregulated at this stage in differentiation, indicating that the cells had progressed past the neural progenitor phase (Fig. 7.15). The retinoic acid interacting genes were also mostly found be downregulated at this stage, returning to levels of expression similar to that of the undifferentiated ESCs. The silencing of pluripotency factors that I observed on day 8 was maintained at this stage (Fig. 7.14) as well as that of *Otx2* and *Fgf5* (Fig. 7.11).

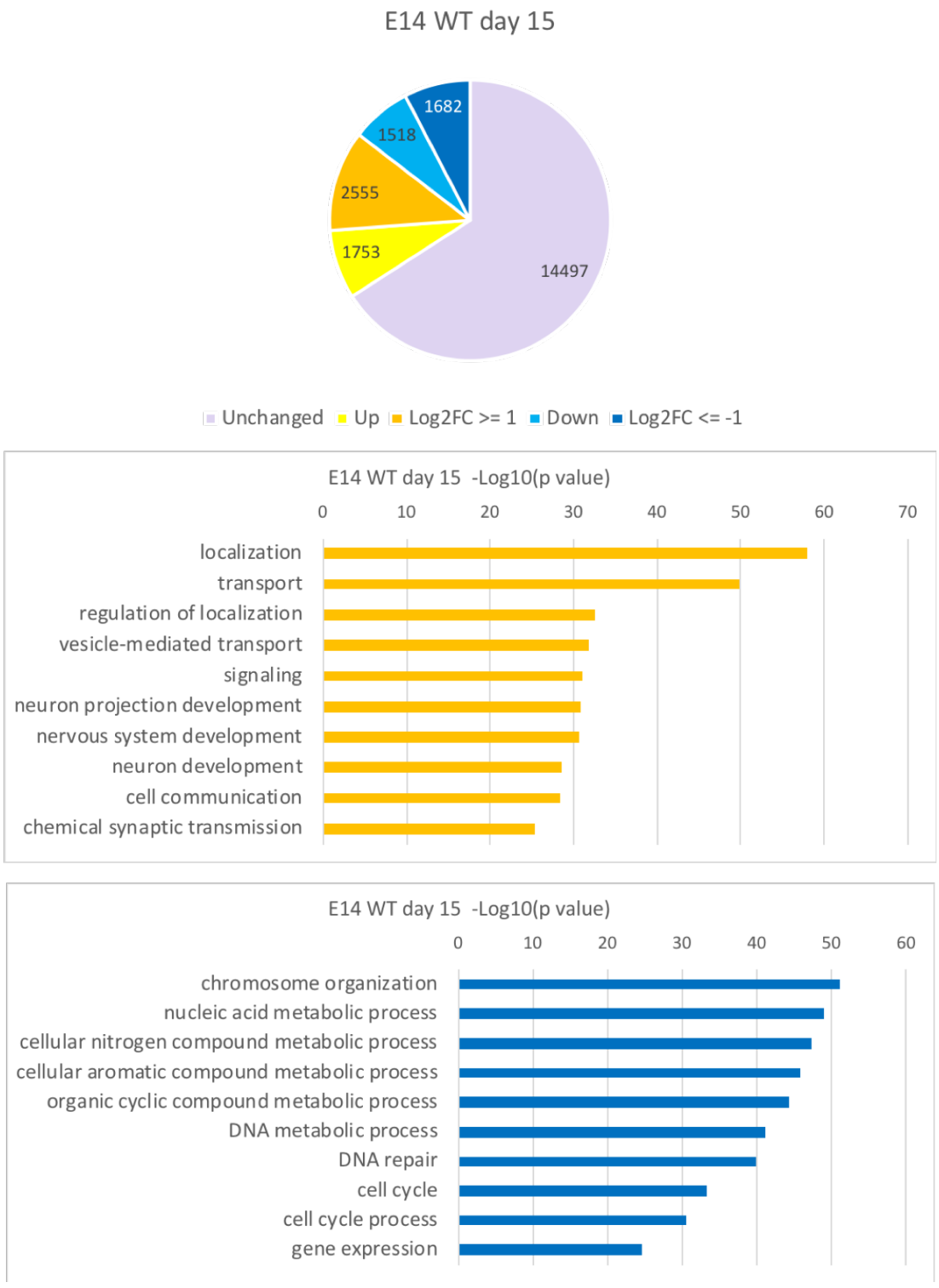


Figure 7.16 Differential gene expression in day 15 E14 control cells relative to day 8. Top panel - Pie chart representing expression of all protein coding genes in day 15 E14 control (E14 WT) cells relative to day 8. Genes significantly upregulated are shown in shades of yellow, genes significantly downregulated are shown in shades of blue. Darker shades correspond to genes differentially expressed by more than a log(2) foldchange of 1 or -1 over day 8. Lower panels - GO term enrichment of significantly upregulated (yellow) and downregulated (blue) genes.

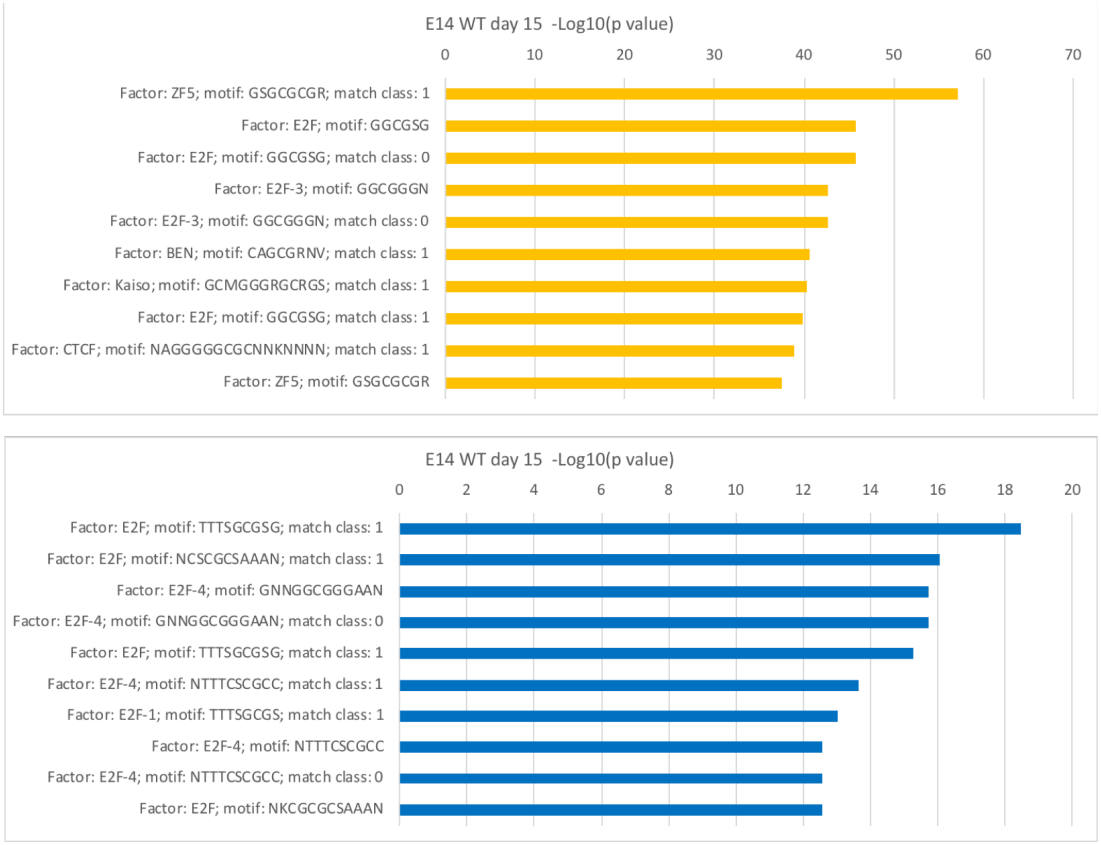


Figure 7.17 Enrichment of transcription factor binding sites in differentially expressed genes in day 15 E14 control. Transcription factor binding motifs enriched in genes significantly upregulated (yellow) and down regulated (blue) in day 15 E14 control (E14 WT) relative to day 8.

Ezh1 and *Ezh2* have been shown to have different expression levels in proliferating and non-proliferating cells, with EZH1 being more highly expressed in adult tissues than EZH2 (Laible et al., 1997; Margueron et al., 2008). I sought to identify the dynamics of expression of the two subunits throughout differentiation. I found that on day 2 and 8, expression of *Ezh2* was maintained at the same level as was seen in the mESCs, but was then significantly downregulated by more than twofold by day 15. *Ezh1* on the other hand was significantly upregulated on day 2 but its expression then decreased again to remain at the same level as was detected in ESCs for the remainder of the timepoints (Fig. 7.19). I also found that *Eed* and *Suz12*, the other essential core subunits of PRC2 were progressively downregulated throughout differentiation. This correlates with the role of EZH1 as being more significant in more differentiated cell types. The considerable decrease in expression of *Eed* and *Suz12* suggests there is an overall lower abundance of the PRC2 complex at these later stages of differentiation.

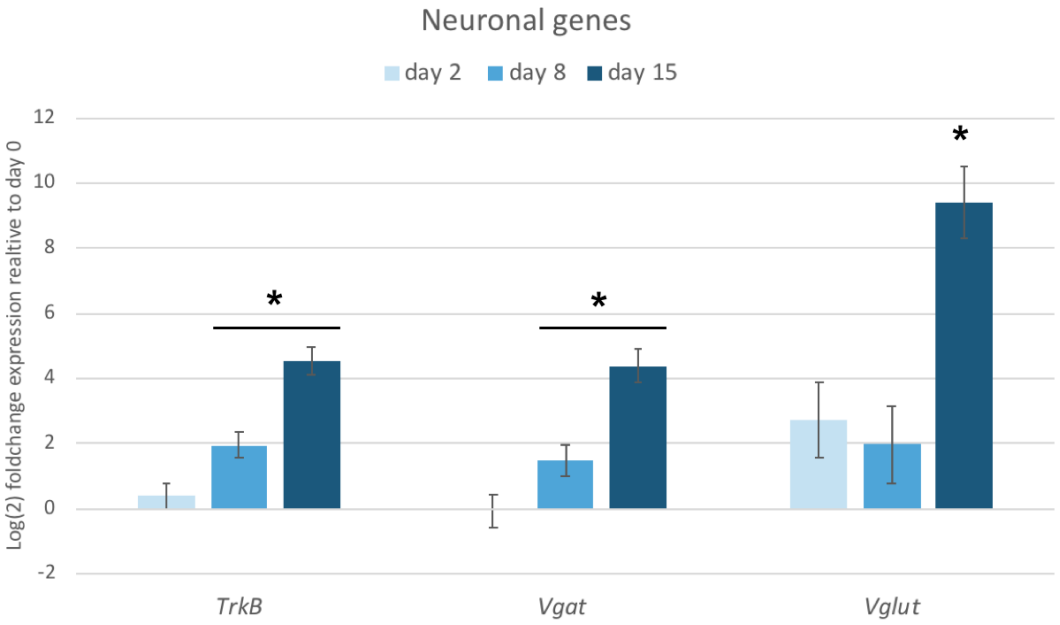


Figure 7.18 Expression of neuronal marker genes in E14 control cells throughout differentiation. Log(2) foldchange of expression of neuronal lineage genes in E14 control at each timepoint relative to undifferentiated E14 control. Significant changes in expression (p value $\leq$0.05) are marked by an asterisk.

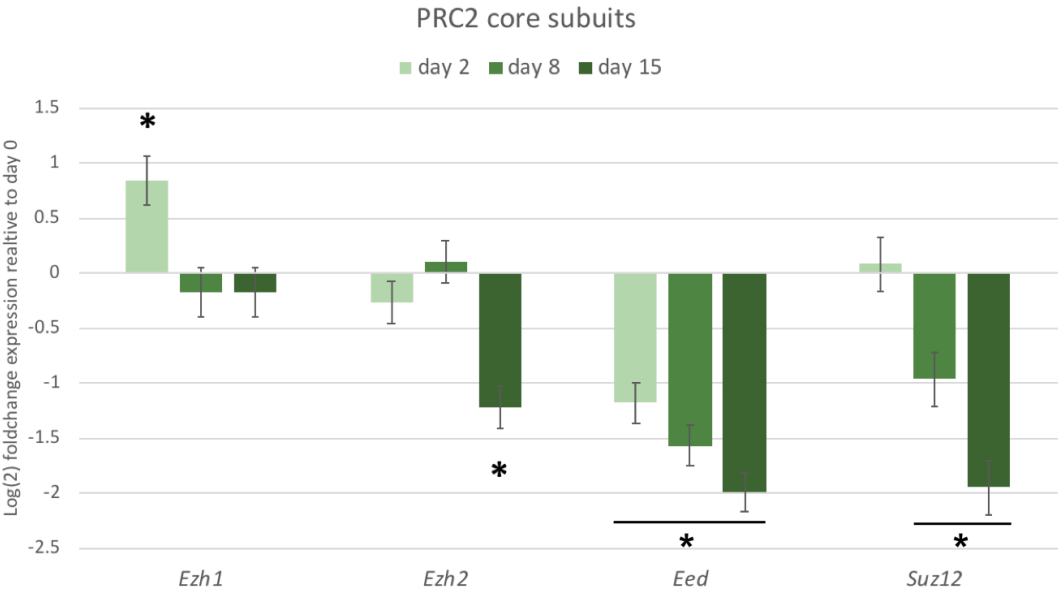


Figure 7.19 Expression of PRC2 subunits in E14 control cells throughout differentiation. Log(2) foldchange of expression of PRC2 core subunits in E14 control cell lines at each timepoint relative to undifferentiated ESCs. Significant changes in expression (p value $\leq$0.05) are marked by an asterisk.

7.5.2 Global differences in gene expression in Polycomb mutant cell lines throughout differentiation

The principle component analysis over all samples showed that the mutants were clustered closely with the E14 control cells at the beginning of differentiation, however, they became progressively more different with each successive timepoint (Fig. 7.20). As expected, I found that throughout differentiation, the mutant background with the largest number of differentially expressed genes was the EZH1/2 dKO cell lines. The EZH2 single mutant lines clustered closely with the E14 control samples until day 2, however, they diverged significantly beginning on day 8. Additionally, I observed for all mutants that there seemed to be a greater sample-to-sample variation than in the E14 control samples. This may reflect an effect caused by the CRISPR editing or may result directly from the loss of PRC2 function and a potential associated decrease in fidelity of maintaining gene expression patterns.

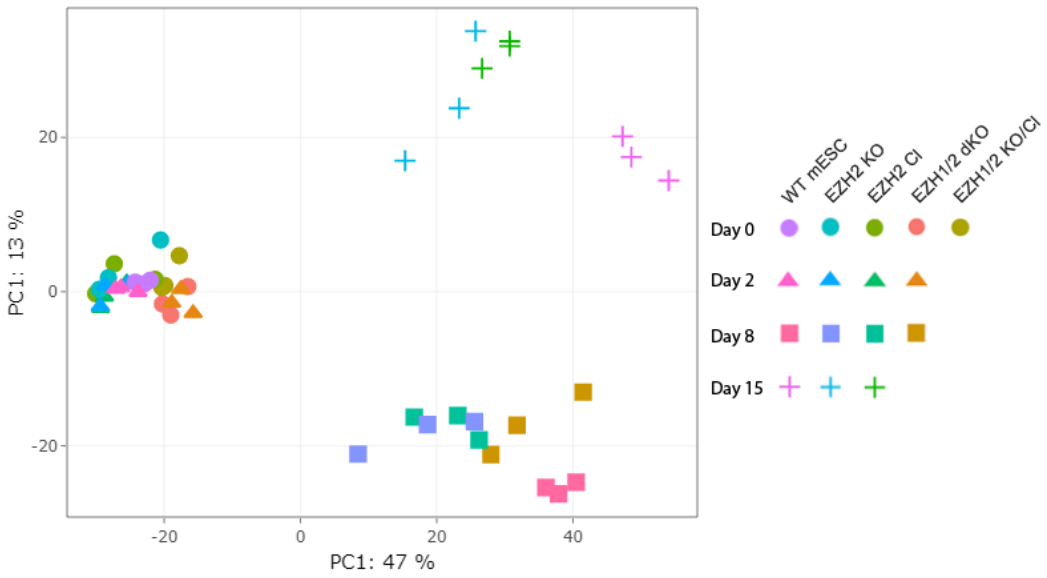


Figure 7.20 Principle component analysis across all RNA-seq samples. E14 control (WT mESC), EZH2 Y726A (EZH2 CI), EZH2 KO, EZH1/2 KO/Y726A (EZH1/2 KO/CI) and EZH1/2 dKO samples were separated into individual replicates. Day 0 samples correspond to undifferentiated ESCs, previously discussed in Chapter 6.

7.5.3 Differential gene expression day 2 EB state

Day 2 corresponds to the second day of EB growth, in which cells have been growing in suspension in the absence of LIF and with 10% serum content for approximately 48 hours. The RNA-seq data from this time point showed that on day 2 there was very little difference in gene expression between the EZH2 single mutant and the E14 control backgrounds, whereas a large number of genes was found to be misregulated for the EZH1/2 dKO background, similar to the situation observed in ESCs.

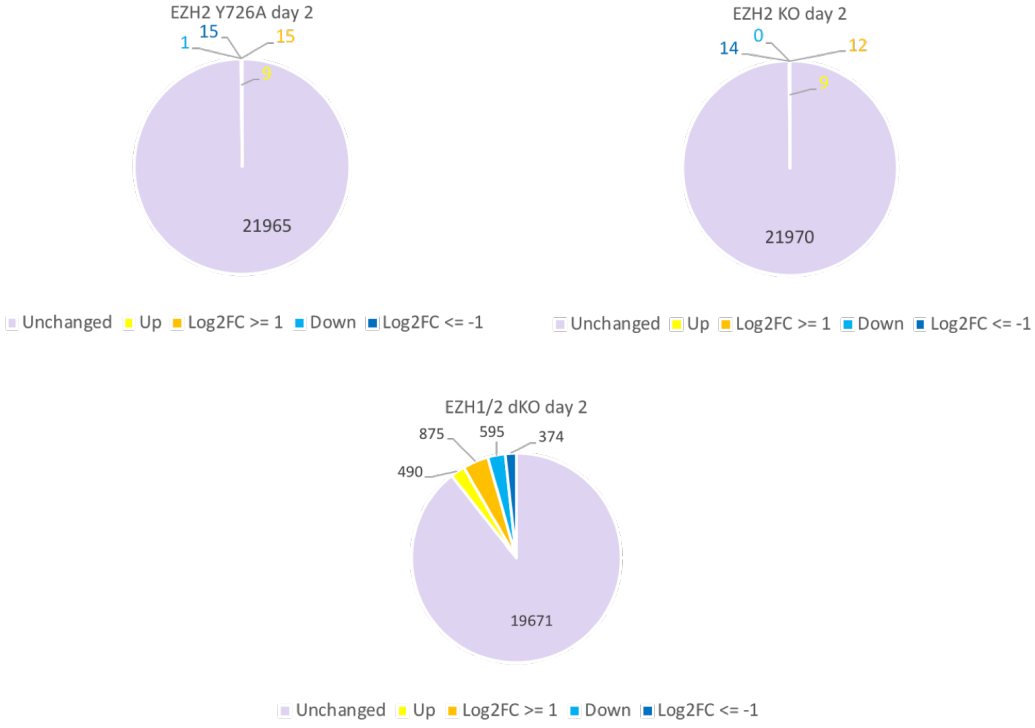


Figure 7.21 Differential gene expression on day 2 in EZH2 single mutant and EZH1/2 dKO cells. Pie charts representing expression of all protein coding genes in day 2 EZH2 Y726A, EZH2 KO and EZH1/2 dKO relative to day 2 E14 control. Genes significantly upregulated are shown in shades of yellow, genes significantly downregulated are shown in shades of blue. Darker shades correspond to genes differentially expressed by more than a log(2) foldchange of 1 or -1 over day 2 E14 control.

Overall changes in gene expression and overlap between mutant background

Consistent with what I observed in ESCs, there were very few genes significantly differentially expressed in the EZH2 single mutants as compared to the E14 control. As can be seen in Fig. 7.21, no more than 40 protein coding genes were significantly affected in either condition. When focusing on the overlap of the genes that did undergo a significant up- or down-regulation in these cell lines I found that approximately one third of these genes were affected in both cell lines (Fig. 7.22). Similar to what was seen for ESCs, a much greater number of genes were differentially expressed in the EZH1/2 dKO cells as compared to the E14 control. More than 2300 protein coding genes underwent significant up- or downregulation in these cells, with more than twice as many genes upregulated by more than twofold as downregulated by this same factor.

GO terms associated with differentially expressed genes

In line with the minor differences in gene expression between the EZH2 single mutant cell lines and the E14 control, there were similarly very few GO terms associated with significantly up- or downregulated genes (Fig. 7.22). In the EZH2 Y726A cells the upregulated genes were enriched for terms associated with senescence associated heterochromatin, as well as the Sgk3 homodimer complex, although this second term was only just over the threshold of significance. These cells were found to downregulate genes associated with 4 different TF classes, 3 of which shared a similar GAAA sequence in their binding motifs. In the EZH2 KO background the Sgk3 homodimer complex was again upregulated with a similar significance as seen in the EZH2 CI, as well as pathways involved in regulating stem cell pluripotency. There were no GO terms found to be associated with any significantly downregulated genes in the EZH2 KO.

In the EZH1/2 dKO cells there was a significant enrichment of GO terms related to embryo development, cell differentiation, and body patterning amongst the upregulated genes, terms strongly represented within Polycomb target genes (Fig. 7.23). Fewer GO terms were found to be associated with the downregulated genes and mostly involved certain developmental abnormalities as well as the DNA polymerase delta complex. In genes both up- and downregulated there was an enrichment of many GC rich transcription factors binding motifs associated with

a variety of transcription factor families (Fig. 7.24).

These findings for the gene expression patterns on day 2 of differentiation strongly correlate with those observed in the undifferentiated mESCs with no increase detected in the number of differences between the PRC2 mutants and the E14 control. This suggests that the contribution of PRC2 to this transition from ESCs to early stage EBs is minimal.

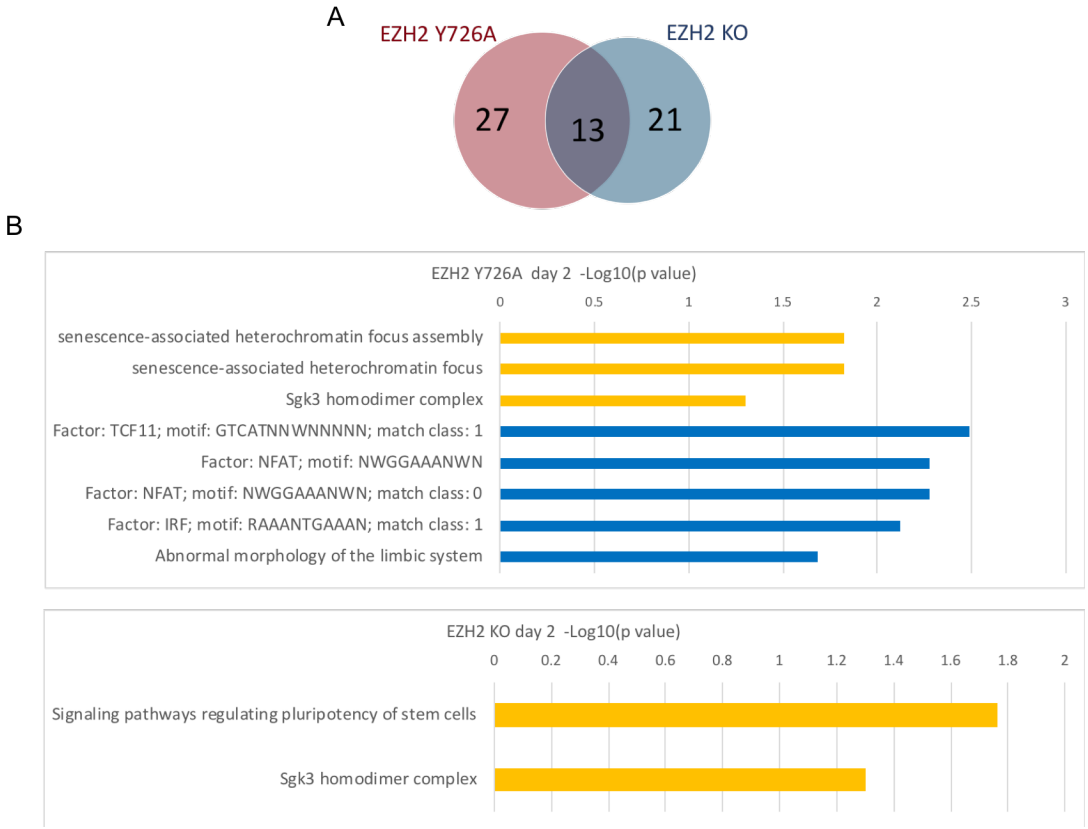


Figure 7.22 Gene ontology of genes differentially expressed on day 2 in EZH2 single mutant cells (**A**)Venn diagram representing overlap of genes differentially expressed in EZH2 Y726A and EZH2 KO. (**B**) GO terms enriched in genes significantly upregulated (yellow) and downregulated (blue) in day 2 EZH2 single mutants relative to day 2 E14 control.

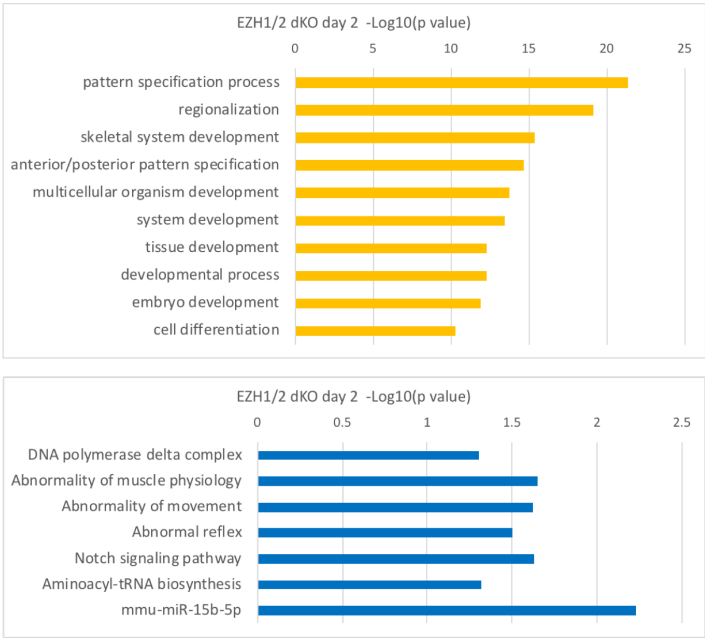


Figure 7.23 Gene ontology of genes differentially expressed on day 2 in EZH1/2 dKO cells. Go terms enriched in genes significantly upregulated (yellow) and downregulated (blue) in day 2 EZH1/2 dKO relative to day 2 E14 control.

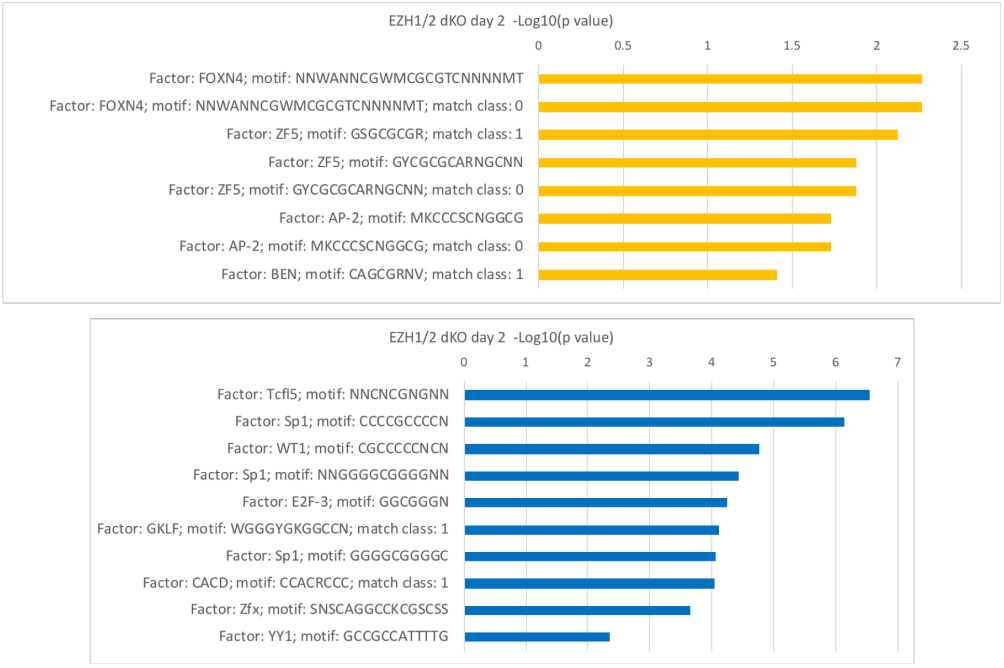


Figure 7.24 Enrichment of transcription factor binding sites in genes differentially expressed on day 2 in EZH1/2 dKO cells. Transcription factor binding motifs enriched in genes significantly upregulated (yellow) and down regulated (blue) in day 2 EZH1/2 dKO relative to day 2 E14 control.

Individual gene expression patterns

When analysing genes known to be affected by withdrawal of LIF I found that in both EZH2 single mutants, the expression of genes either upregulated or downregulated in E14 control cells was not significantly altered, even though *Klf4* and *Tbx3* were expressed to higher levels albeit not above the threshold of significance (Fig. 7.25). In the EZH1/2 dKO background, however, there were some differences observed. Although both *Otx2* and *Fgf5* were found to be expressed at similar levels to what was observed in the E14 control, *Klf4* or *Tbx3* were both upregulated in this mutant compared to the E14 control background, indicating that these genes were not appropriately silenced in response to LIF withdrawal. I found that these genes were also overexpressed in the undifferentiated EZH1/2 dKO ESCs as compared to the E14 control. Interestingly, when comparing expression at these 2 timepoints I noticed that expression of *Klf4* and *Tbx3* did actually decrease upon withdrawal of LIF in the EZH1/2 dKO cell lines, however, repression was less effective than in the E14 control cells.

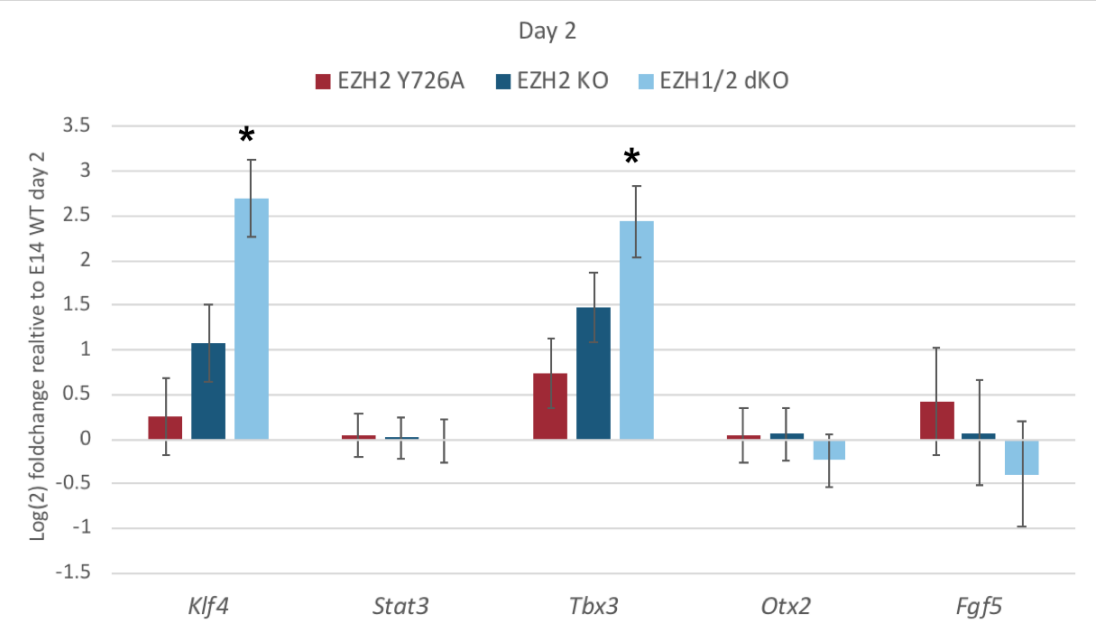


Figure 7.25 Expression of developmental genes in PRC2 mutants on day 2 of differentiation. Log(2) foldchange of expression of genes known to be affected by withdrawal of LIF in EZH2 Y726A (EHZ2 CI), EZH2 KO and EZH1/2 dKO relative to E14 control on day 2. Significant changes in expression (p value <0.05) are marked by an asterisk.

Overall, with the exception of *Klf4* and *Tbx3*, expression of this subset of genes known to display dynamic expression in response to the withdrawal of LIF was similar between the mutants backgrounds and E14 control cells. However it did appear that there was some slight disruption of this process in the EZH1/2 dKO cells, specifically in silencing of genes.

7.5.4 Differential gene expression day 8 transition to neural progenitor state

Day 8 corresponded, in the E14 control, to the stage at which cells had become *Pax6* positive neural progenitors, more specifically, of the radial glial type as denoted by their expression of *Glast*, *Fabp7*, and *Nestin*. The principle component analysis showed that this was the timepoint at which the EZH2 single mutant cell lines began to strongly diverge from the E14 control cells. I therefore analysed changes in expression at this time point in greater detail.

Overall changes in gene expression and overlap between mutant backgrounds

By day 8 the number of differentially expressed genes in the EZH2 single mutant cells was greatly increased as compared to what was observed on day 2 (Fig. 7.26). The EZH2 Y726A cells were found to have more than 2300 protein coding genes differentially expressed of which nearly 1000 genes were upregulated and just over 1300 were downregulated. In the EZH2 KO cell lines these numbers were almost doubled with more than 4000 protein coding genes differentially expressed of which nearly 2000 genes were upregulated.

Most of the genes affected in the EZH2 Y726A background were found to be similarly misregulated in the EZH2 KO, however around half of the genes differentially expressed in the KO were not found to be significantly altered in the EZH2 Y726A background compared to the E14 control cells at that stage (Fig. 7.27). In light of this considerable difference, I decided to compare gene expression in the two mutants to each other to reveal genes differentially expressed between EZH2 Y726A and EZH2 KO. Surprisingly however, I found that, despite the apparent large difference between the 2 mutants, only 6 genes were significantly differentially expressed between them, one of which was EZH2 itself (Fig. 7.27).

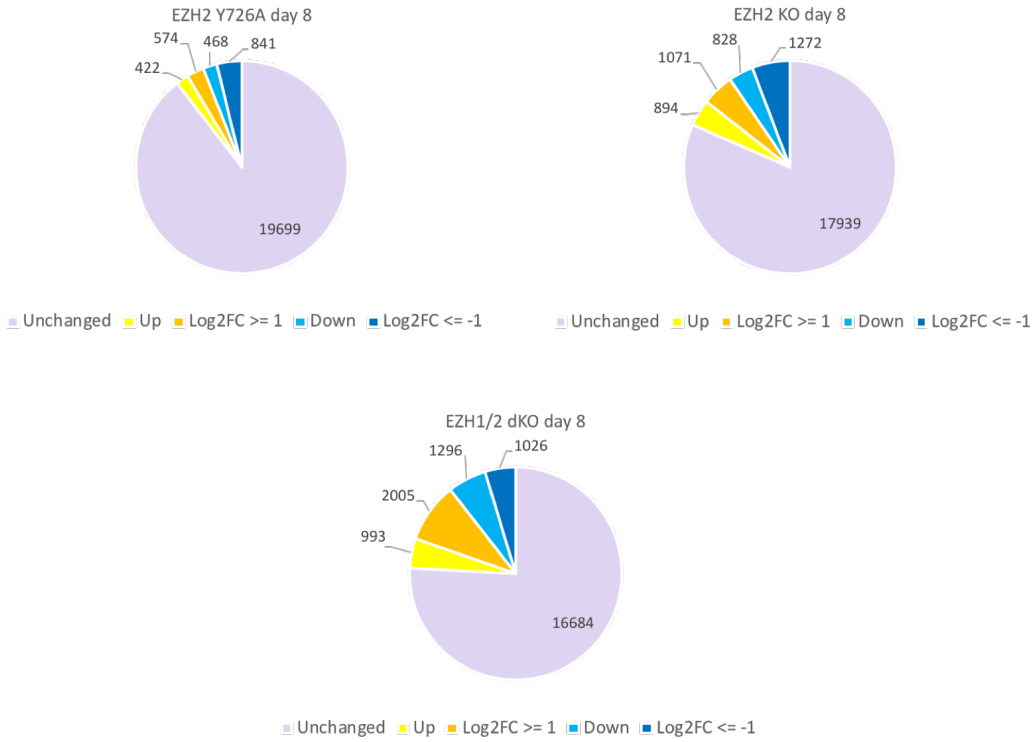
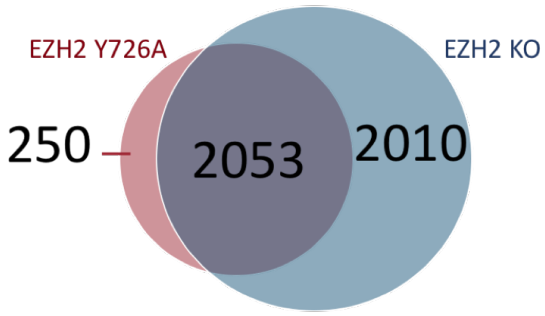


Figure 7.26 Differential gene expression on day 8 in EZH2 single mutant and EZH1/2 dKO cell lines. Pie charts representing expression of all protein coding genes in day 8 EZH2 Y726A, EZH2 KO and EZH1/2 dKO relative to day 8 E14 control. Genes significantly upregulated are shown in shades of yellow, genes significantly downregulated are shown in shades of blue. Darker shades correspond to genes differentially expressed by more than a log(2) foldchange of 1 or -1 over day 8 E14 control cells.

Pax2, a transcription factor involved in development of the CNS amongst other systems, was also downregulated in the EZH2 KO cells. I hypothesised that this very low number of significantly differentially expressed genes might indicate that the genes that were significantly misregulated compared to the E14 control in the EZH2 KO background only may still follow a similar trend in expression in the EZH2 Y762A cell lines and vice versa, but the significance threshold was only reached in one or the other of them.

A



B

Gene	log2FoldChange	padj	threshold
Ece1	-0.886	0.04034	adjp
Pax2	-3.581	0.00296	both
Ezh2	-1.176	7.7E-06	both
Frem1	-2.710	0.0148	both
Gm21874	-34.985	0.00041	both
FMN1	-22.297	0.00047	both

Figure 7.27 Comparison of gene expression between EZH2 Y726A and EZH2 KO cells. (**A.**) Venn diagram representing overlap of genes differentially expressed in day 8 EZH2 Y726A and EZH2 KO compared to E14 control cells. (**B.**) Table of all genes found to be differentially expressed between EZH2 Y726A and EZH2 KO cell lines at day 8. Adjusted p values are labelled as padj, threshold column indicates if p value is below 0.05 and if log(2) foldchange is greater than 1 or below -1.

With over 5300 genes, or nearly a quarter of all protein coding genes, misregulated as compared to the E14 control, the EZH1/2 dKO was again found to have more genes differentially expressed than either of the 2 EZH2 single mutant cell lines (Fig. 7.26). Overall, more than 2300 genes were found to be significantly downregulated while nearly 3000 were upregulated. Of the genes significantly upregulated, more than 2000 were increased in expression by more than twofold, nearly twice as many as the 1026 genes found to decrease in expression by the same factor. These large-scale differences in gene expression reflect the considerable difference in EB morphology and divergence of these cells from the E14 control at this stage of differentiation. As expected following the depletion of a repressive complex, most misregulated genes were upregulated, incidentally, this number was close to the total number of Polycomb bound genes identified in ESCs (Mohn et al., 2008). More surprising was the number of downregulated genes, some of which may be repressed as an indirect consequence of the upregulation of Polycomb targeted repressors. Other downregulated genes may simply fail to be upregulated in a PRC2 depleted context.

GO terms associated with differentially expressed genes

Both EZH2 single mutant cell lines displayed an enrichment for GO terms associated with spermatogenesis and meiosis amongst the genes that were upregulated. Additionally, in the EZH2 Y726A there was an upregulation of factors involved in the immune response, while in the EZH2 KO there was an upregulation of genes associated with the response to LIF. Amongst the downregulated genes in both EZH2 single mutant backgrounds the GO terms most strikingly enriched were those associated with neuronal development and cell differentiation, indicating that they failed to be correctly upregulated as compared to the E14 control cells (Fig. 7.28 and 7.29).

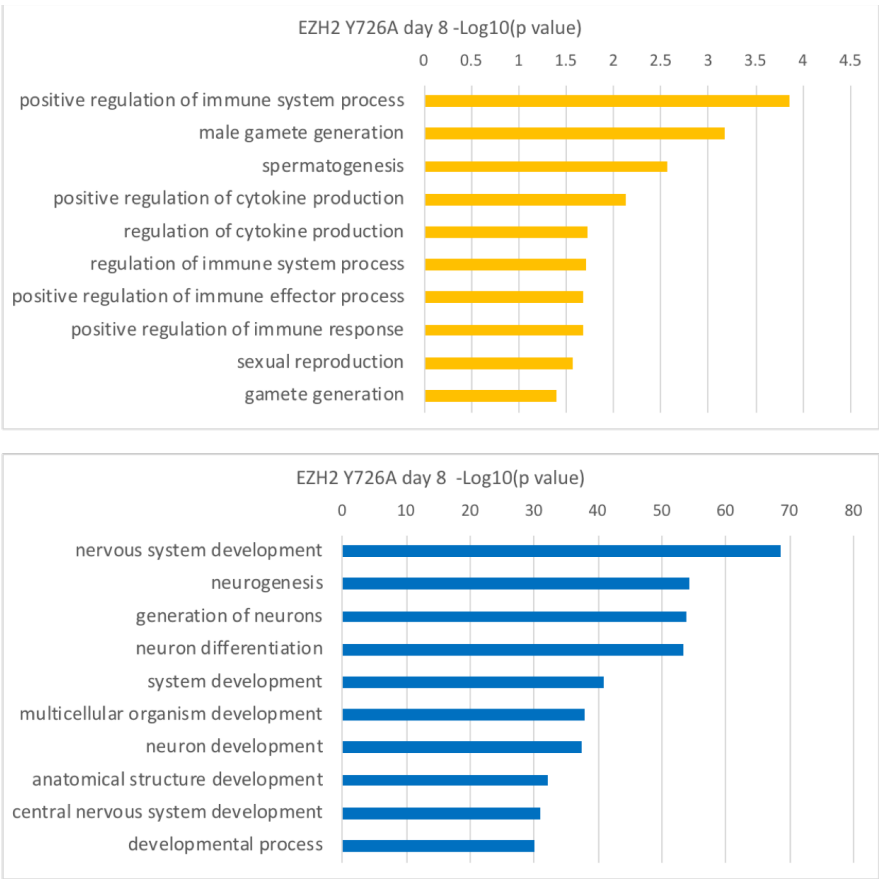


Figure 7.28 Gene ontology of genes differentially expressed on day 8 in EZH2 Y726A cells. Go terms enriched in genes significantly upregulated (yellow) and downregulated (blue) in day 8 EZH2 Y726A cells relative to day 8 E14 control.

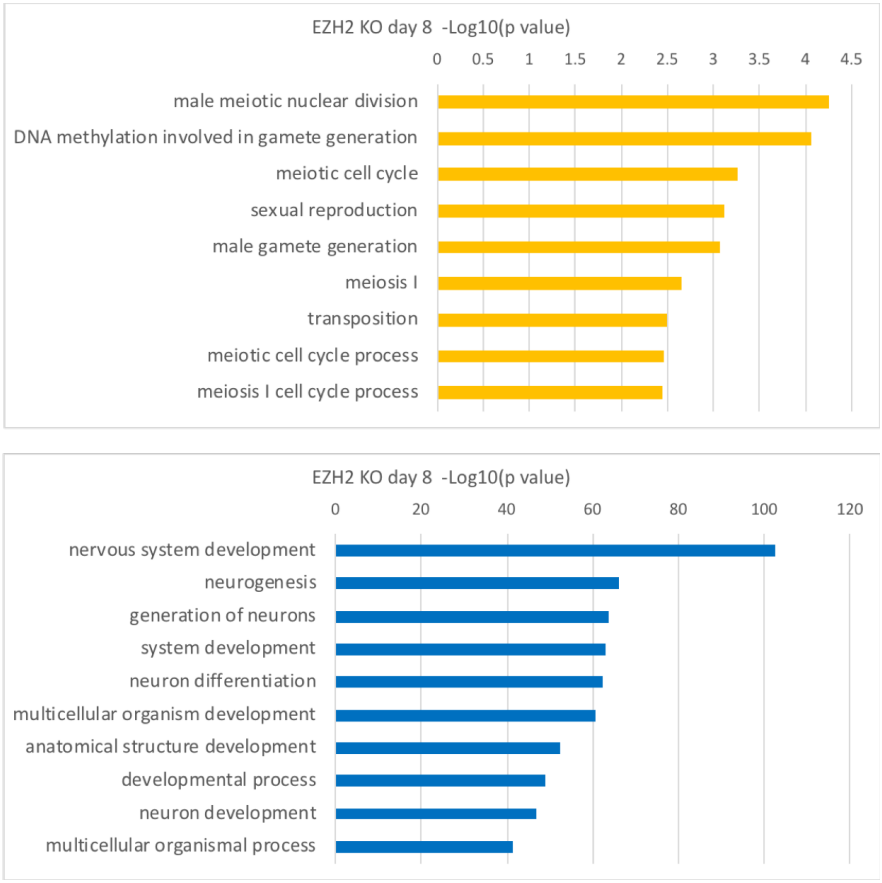


Figure 7.29 Gene ontology of genes differentially expressed on day 8 in EZH2 KO cells. Go terms enriched in genes significantly upregulated (yellow) and downregulated (blue) in day 8 EZH2 KO cells relative to day 8 E14 control.

Only 2 transcription factor binding motifs were enriched in the genes upregulated in the EZH2 Y726A, but there were many more found in those upregulated in the EZH2 KO background. Of these, many belonged to transcription factors known to be Polycomb target genes such as *Otx2*, *Pax6* and *Hoxa13*, or retinoic acid nuclear receptors such as RXR-alpha and RXR-gamma (Fig. 7.30 and 7.31). The enrichments of TF binding motifs in the genes downregulated in both EZH2 single mutant backgrounds were very numerous and often CG rich sequences, and many were identified in both cell lines. Of note were motifs for GKLF, Kaiso, ZF5, RNF96 and MOVO-B all of which were found to to be enriched in the upregulated genes for the E14 control cells on day 8, further demonstrating a lack of upregulation of neuronal factors in the EZH2 single mutant cells at this timepoint.

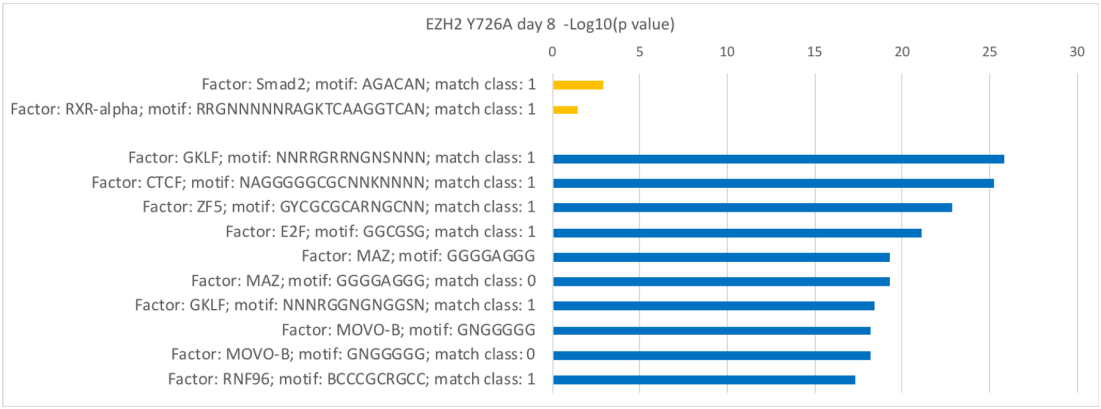


Figure 7.30 Enrichment of transcription factor binding sites in genes differentially expressed on day 8 in EZH2 Y726A cells. Transcription factor binding motifs enriched in genes significantly upregulated (yellow) and down regulated (blue) in day 8 EZH2 Y726A cells relative to day 8 E14 control.

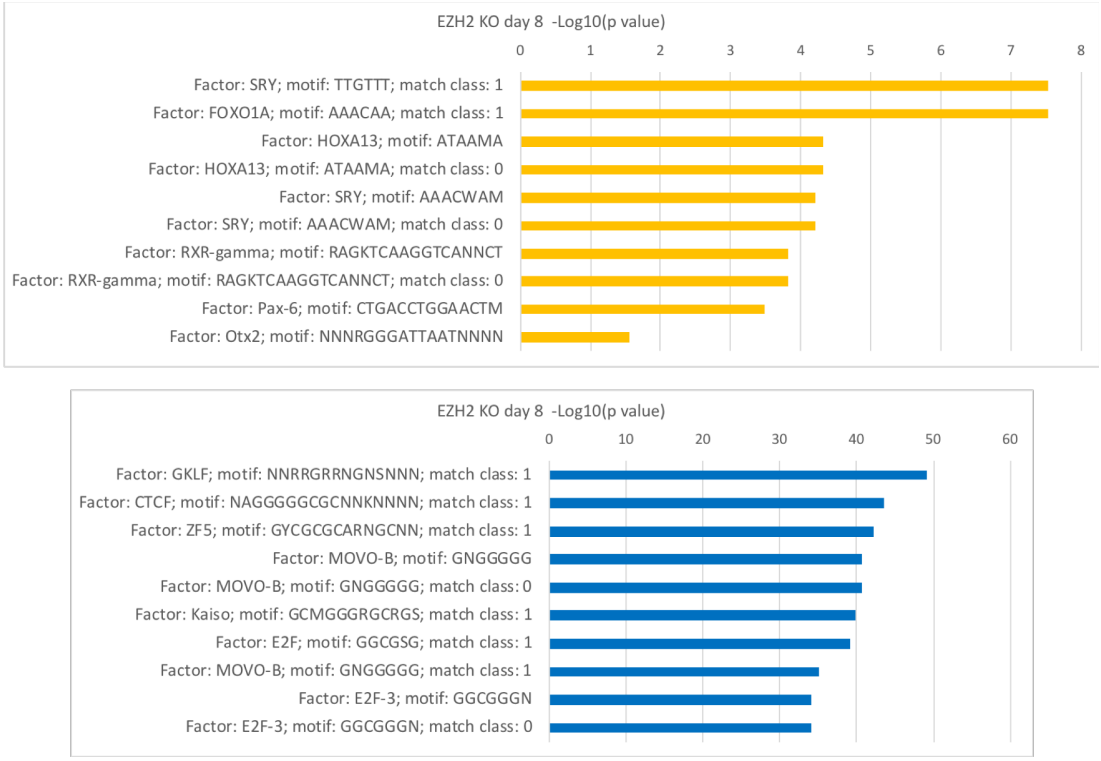


Figure 7.31 Enrichment of transcription factor binding sites in genes differentially expressed on day 8 in EZH2 KO cells. Transcription factor binding motifs enriched in genes significantly upregulated (yellow) and down regulated (blue) in day 8 EZH2 KO cells relative to day 8 E14 control.

The EZH1/2 dKO cells also saw an upregulation of factors associated with the immune response as well as transmembrane transport, while similarly experiencing a downregulation of genes involved in the development of the nervous

system (Fig. 7.32). In addition, there was a downregulation of genes associated with certain metabolic systems in these cells as well as certain factors involved in DNA packaging and chromosome organisation, suggesting a slowdown in rates of cell division. As was seen in the single mutants at this time point, there was an enrichment in downregulated genes of binding motifs for ZF5 which is known to positively regulate neuronal development. FOXN4, a transcription factor involved in cell fate decisions in the developing CNS amongst others (Xiang and Li, 2013), also saw an enrichment of its binding site amongst downregulated genes (Fig. 7.33). As with the EZH2 single mutant cells at this timepoint, there was a general failure to upregulate neuronal genes in the EZH1/2 dKO on day 8. The upregulation of genes associated with the immune response may reflect the overall decreased viability of these cells at this stage of differentiation as they displayed a noticeably higher rate of cell death even by day 6 (see Fig. 7.2).

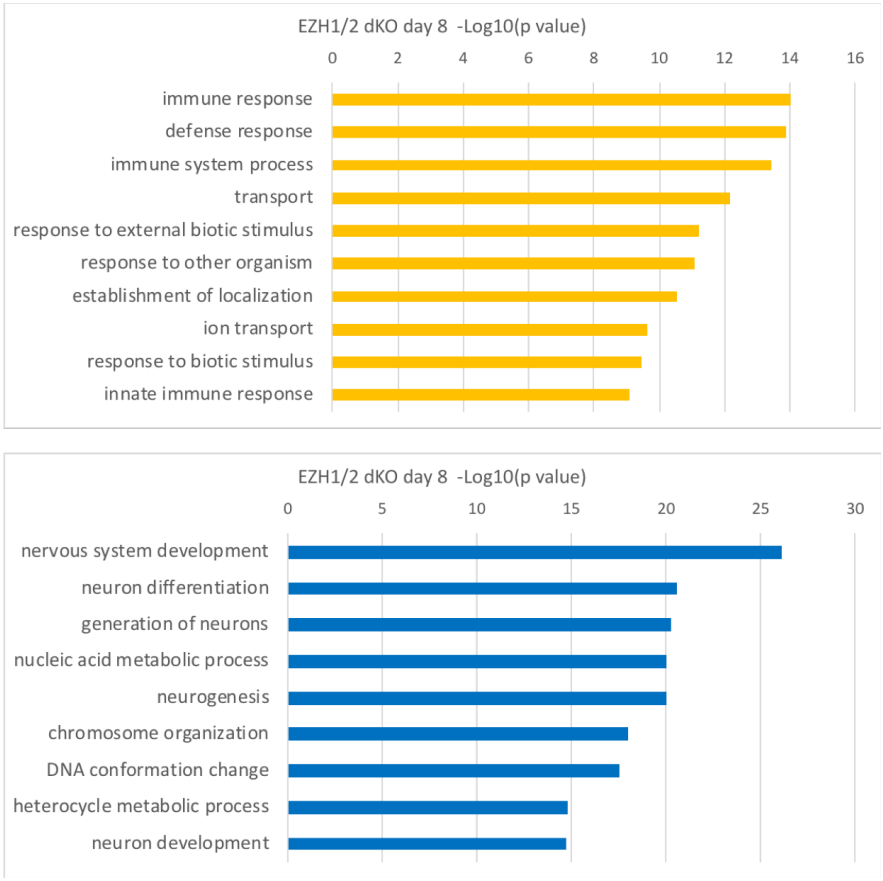


Figure 7.32 Gene ontology of genes differentially expressed on day 8 in EZH1/2 dKO cells. Go terms enriched in genes significantly upregulated (yellow) and downregulated (blue) in day 8 EZH1/2 dKO relative to day 8 E14 control.

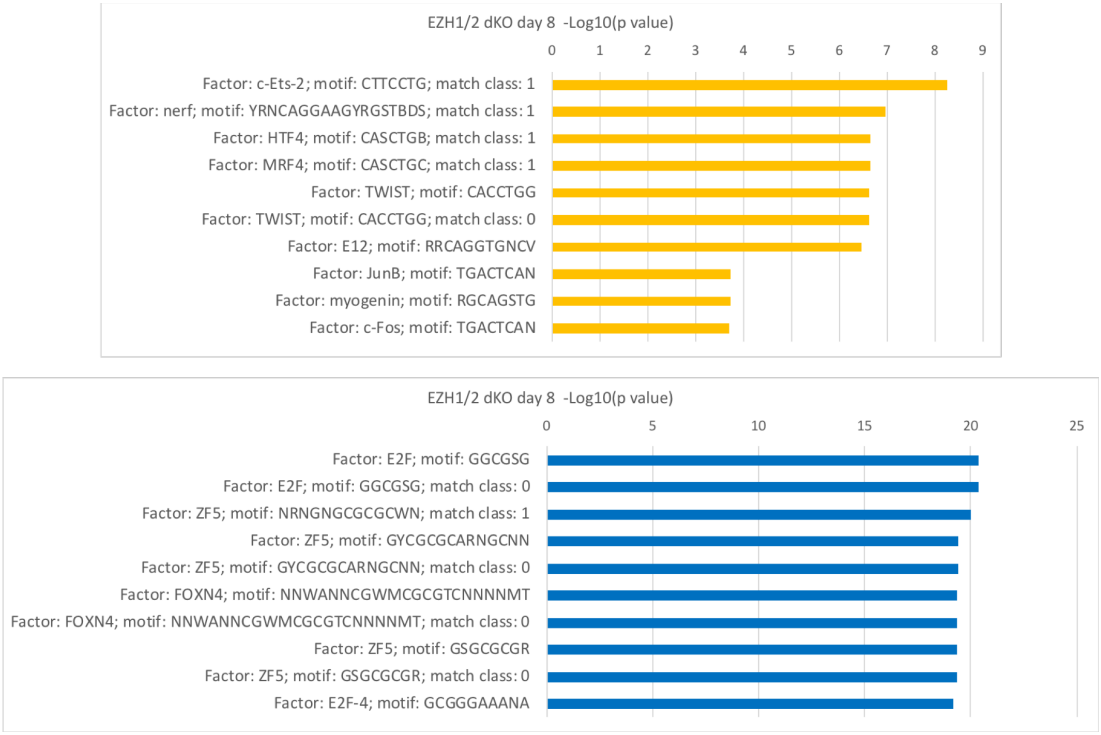


Figure 7.33 Enrichment of transcription factor binding sites in genes differentially expressed on day 8 in EZH1/2 dKO cells. Transcription factor binding motifs enriched in genes significantly upregulated (yellow) and down regulated (blue) in day 8 EZH1/2 dKO relative to day 8 E14 control.

Individual gene expression patterns

Of the markers of neural precursor cell identity mentioned above, I found that in the EZH2 single mutants, none were upregulated to the extent that was seen in the E14 control cells (Fig. 7.37). Of the pluripotency factors assessed I found that both *Oct4* and *Nanog* failed to be properly downregulated in both EZH2 single mutant cells, however *Sox2* was repressed as was seen in the E14 control (Fig. 7.34). Despite this significant difference in the expression of *Oct4* and *Nanog* in the EZH2 single mutants relative to the E14 control on day 8 however, as compared to the levels detected in ESCs (day 0) the genes were both considerably downregulated (Fig. 7.35 and 7.36). Both *Fgf5* and *Otx2* were downregulated in the EZH2 single mutants however *Otx2* was still expressed to a level that was significantly higher than that found in the E14 control, as was also the case for *Tbx3* in both mutants (Fig. 7.38). Of the retinoic acid transporters and nuclear receptors I found that some but not all failed to undergo the induction of expression observed in the E14 control cells, as some were significantly underexpressed in the EZH2 single mutants (Fig. 7.37).

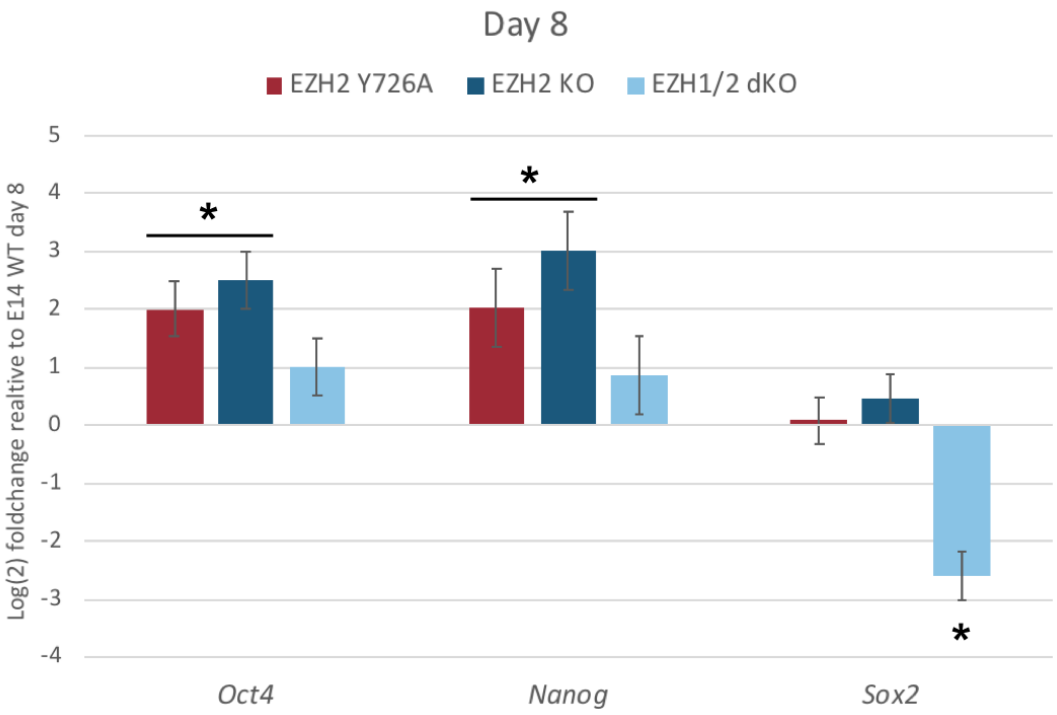


Figure 7.34 Expression of pluripotency factors in PRC2 mutant cells on day 8 of differentiation. Log(2) foldchange of expression of pluripotency factors in EZH2 Y726A, EZH2 KO and EZH1/2 dKO relative to E14 control on day 8. Significant changes in expression (p value ≤ 0.05) are marked by an asterisk.

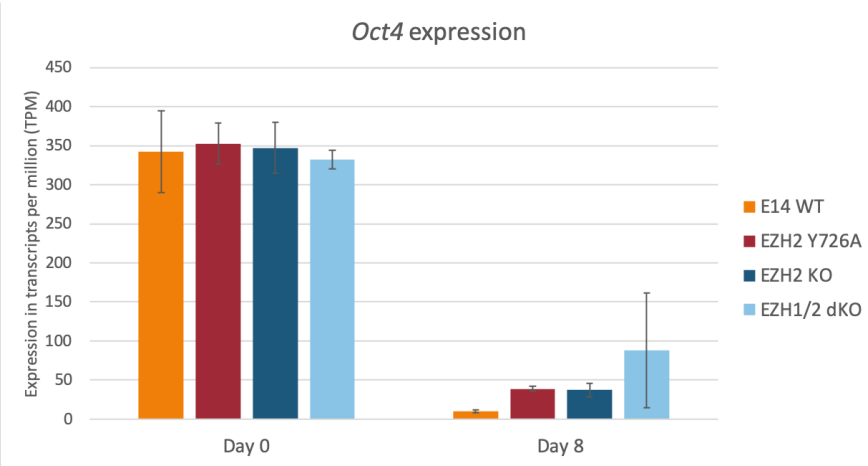


Figure 7.35 Expression dynamics of pluripotency factor *Oct4* between day 0 and day 8. Average number of transcripts per million detected for *Oct4* across the three replicates of E14 control (E14 WT), EZH2 Y726A, EZH2 KO and EZH1/2 dKO on day 0 and day 8. Error bars represent SEM across the three replicates.

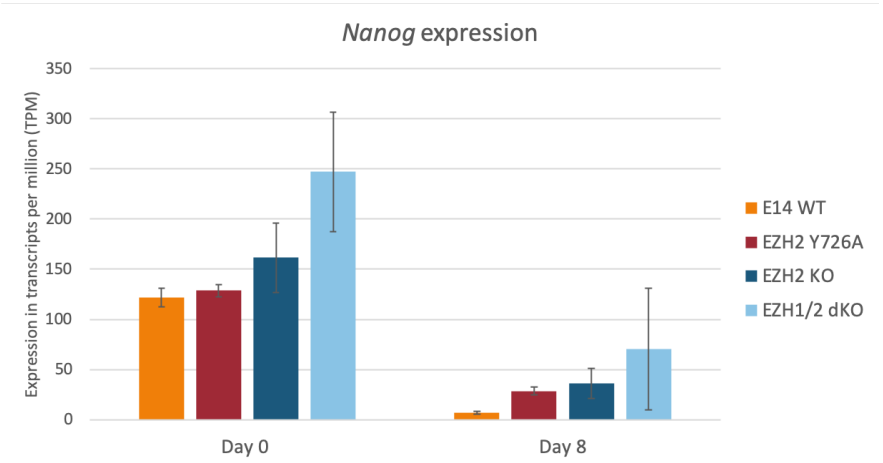


Figure 7.36 Expression dynamics of pluripotency factor *Nanog* between day 0 and day 8. Average number of transcripts per million detected for *Nanog* across the three replicates of E14 control (E14 WT), EZH2 Y726A, EZH2 KO and EZH1/2 dKO on day 0 and day 8. Error bars represent SEM across the three replicates.

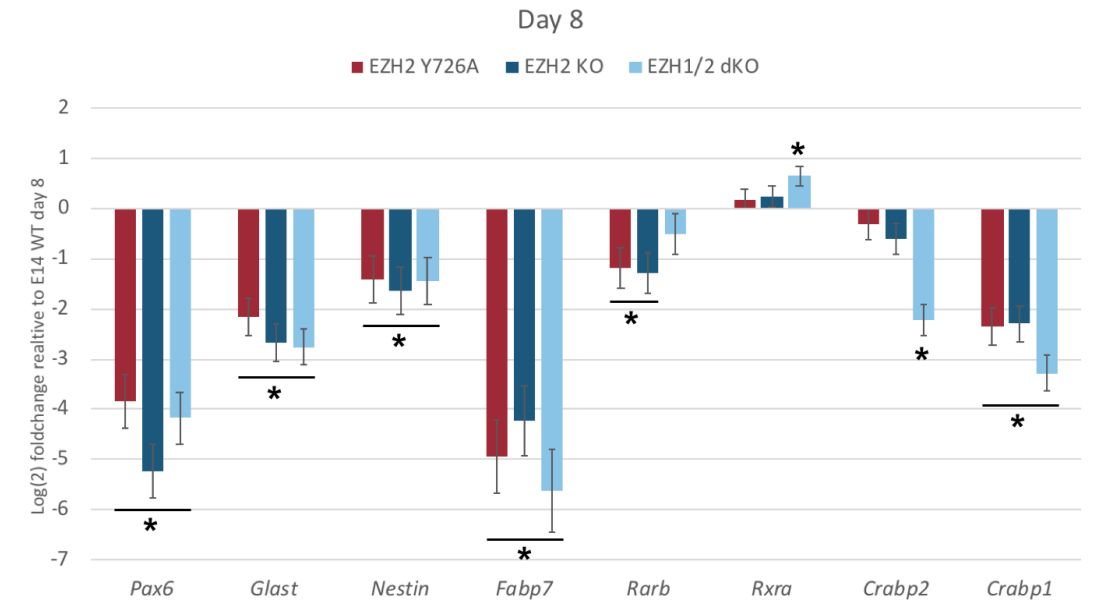


Figure 7.37 Expression of NPC markers and retinoic acid interactors in PRC2 mutant cells on day 8 of differentiation. Log(2) foldchange of expression of neural progenitor marker genes and retinoic acid interacting genes in EZH2 Y726A, EZH2 KO and EZH1/2 dKO relative to E14 control on day 8. Significant changes in expression (p value ≤ 0.05) are marked by an asterisk

In the EZH1/2 dKO cells, although I found no upregulation of the markers of the neural progenitor phenotype, I did observe that pluripotency factors as well as *Fgf5* were silenced to similar levels as were seen in the E14 control (Fig. 7.37 and

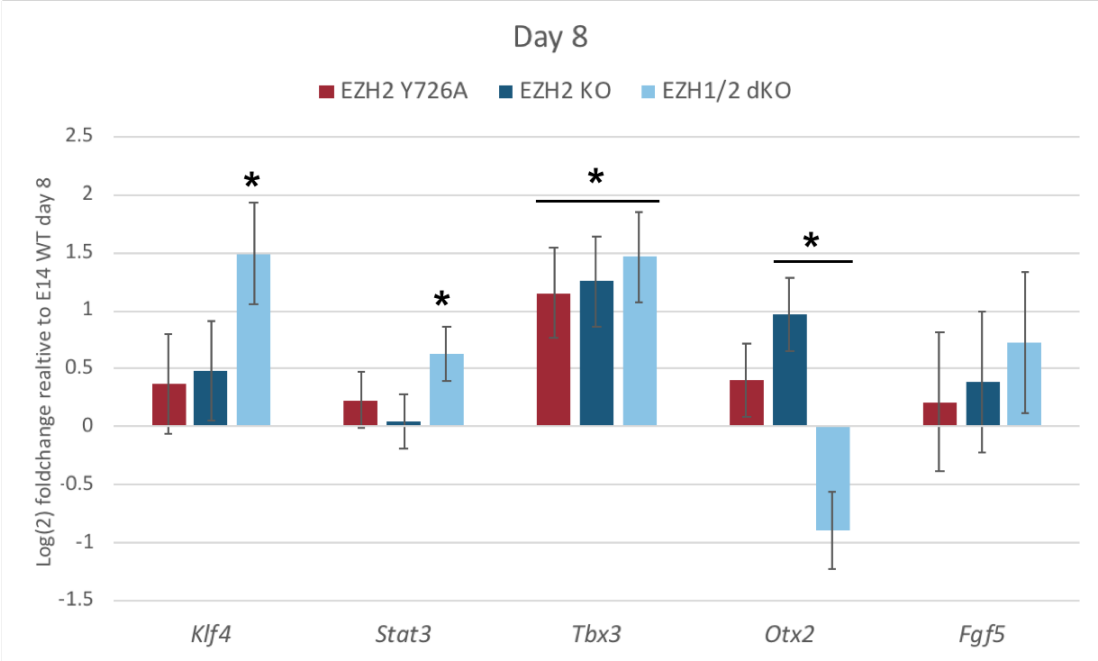


Figure 7.38 Expression of developmental genes in PRC2 mutant cells on day 8 of differentiation. Log(2) foldchange of expression of genes known to be affected by withdrawal of LIF in EZH2 Y726A, EZH2 KO and EZH1/2 dKO relative to E14 control on day 8. Significant changes in expression (p value $\leq$0.05) are marked by an asterisk.

7.38). Although there was no significant difference in the levels of *Oct4* and *Nanog* detected in the EZH1/2 dKO as compared to the E14 control at this timepoint (Fig. 7.34) there was a higher average number of transcripts detected for both genes in this cell line, suggesting a high level of variability between samples here (Fig. 7.35 and 7.36) Similarly to what I observed in the EZH2 single mutants, I found that in the EZH1/2 dKO the retinoic acid binders were not all upregulated as they were in the E14 control (Fig. 7.37).

Hox gene expression was found to go through a drastic change between day 2 and day 8 in the E14 control cells, with the majority of Hox genes becoming upregulated on day 8 (Fig. 7.39A). I found that none of the Hox genes for which silencing was maintained in the E14 control were significantly overexpressed in either of the EZH2 single mutant backgrounds on day 8 (Fig. 7.39B). However, almost all of these genes were derepressed in the EZH1/2 dKO (Fig. 7.39). Of the Hox genes significantly upregulated in the E14 control, those with only modest increases in expression compared to day 2 were not significantly differentially expressed in either EZH2 single mutant background on day 8, but their expression was significantly higher in the EZH1/2 dKO background. For the Hox genes which

saw a strong induction in the E14 control from day 2 to day 8, however, I found that both in the EZH2 single mutants and the EZH1/2 dKO the expression of almost all of these genes was significantly lower. These results suggest that while the EZH2 single mutants were able to maintain a certain level of gene repression that the EZH1/2 dKO failed to achieve, all 3 mutants were lacking an activating factor that is required for the Hox genes to reach the level of expression seen in the E14 control on day 8.

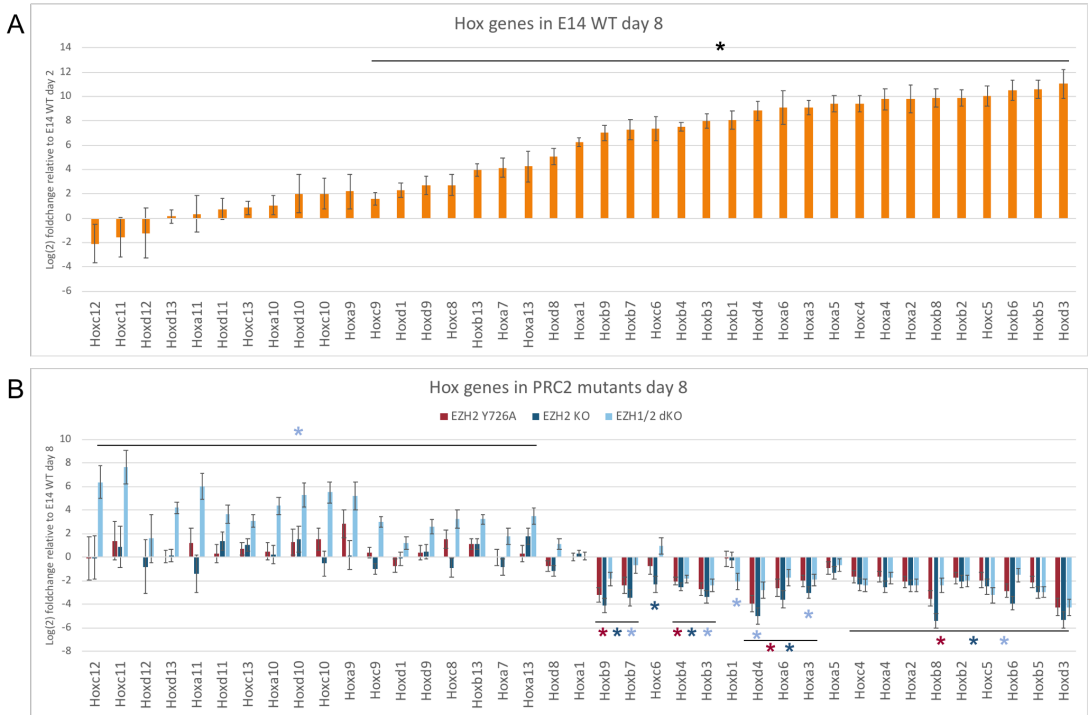


Figure 7.39 Expression of Hox genes on day 8 of differentiation. (**A.**) Log(2) foldchange of expression of Hox genes in E14 control on day 8 relative to day 2 E14 control. Significantly upregulated genes (p value $\leq$0.05) are marked by an asterisk. (**B.**) Log(2) foldchange of expression of Hox genes in EZH2 Y726A, EZH2 KO and EZH1/2 dKO on day 8 relative to day 8 E14 control. Significantly upregulated genes (p value $\leq$0.05) are marked by an asterisk of the colour corresponding to each mutant cell line.

7.5.5 Day 15 differential gene expression

In the E14 control on day 15 cells had progressed through differentiation to a more mature neuronal phenotype, expressing factors such ad *TrkB*, *Vgat* and *Vglut*, and downregualting *Pax6*. As on day 8, the EZH2 single mutants were considerably divergent from the E14 control. As previously mentioned, by day 15 there were no surviving EZH1/2 dKO cells on the plates.

Overall changes in gene expression and overlap between mutants

On day 15 both EZH2 single mutants were found to misregulate an even greater number of genes than on day 8. The EZH2 Y726A cells were found to differentially express more than 5000 protein coding genes of which a similar number were found to be upregulated as were downregulated. The EZH2 KO cells were again found to differentially express more genes than the EZH2 Y726A at almost 6300 protein coding genes of which there were around 400 more genes upregulated than downregulated. In both mutants there were more genes downregulated by more than twofold than there were genes upregulated by this same factor (Fig. 7.40).

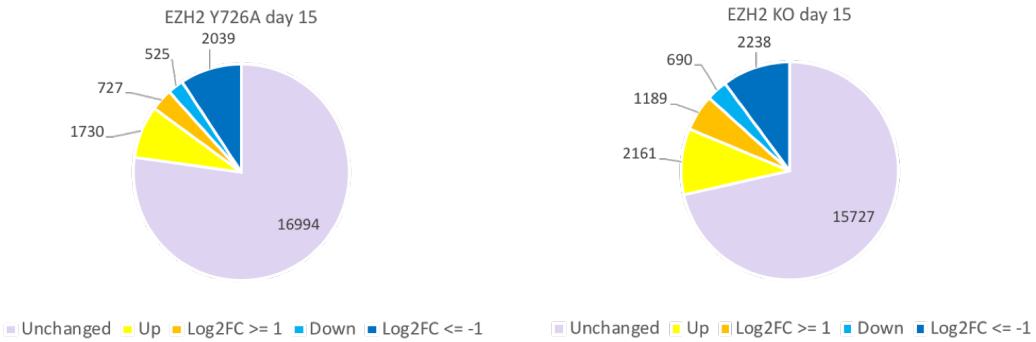


Figure 7.40 Differential gene expression in day 15 EZH2 single mutants. Pie charts representing expression of all protein coding genes in day 15 EZH2 Y726A and EZH2 KO relative to day 15 E14 control. Genes significantly upregulated are shown in shades of yellow, genes significantly downregulated are shown in shades of blue. Darker shades correspond to genes differentially expressed by more than a log(2) foldchange of 1 or -1 over day 15 E14 control.

Nearly 4000 genes were found to be differentially expressed in both EZH2 KO and EZH2 Y726A as compared to the E14 control with just over 1000 genes only affected in the EZH2 Y726A and more than twice as many only affected in the EZH2 KO (Fig. 7.41A). When I compared gene expression between the 2 mutant cell lines at this timepoint, I found there to be just under 300 genes differentially expressed. Of these genes nearly 200 were found to be upregulated in the EZH2 KO as compared to the EZH2 Y726A, while just under 100 were downregulated (Fig. 7.41B). This was a greater difference than was observed between each of the cell lines on day 8 (see Fig. 7.27) however compared to the number of genes differentially expressed as compared to the E14 control, this was still relatively low. This suggests that by day 15 the EZH2 KO and EZH2 Y726A cells were becoming more divergent from each other than they had been throughout the earlier stages of differentiation.

Figure 7.41 Comparison of gene expression between EZH2 Y726A and EZH2 KO on day 15. (**A.**) Venn diagram representing overlap of genes differentially expressed in day 15 EZH2 Y726A and EZH2 KO. (**B.**) Pie chart representing expression of all protein coding genes in EZH2 KO relative to EZH2 Y726A on day 15. Genes significantly upregulated in EZH2 KO are shown in shades of yellow, genes significantly downregulated are shown in shades of blue. Darker shades correspond to genes differentially expressed by more than a log(2) foldchange of 1 or -1 over EZH2 Y726A. (**C.**) GO terms enriched in genes significantly up (yellow) and downregulated (blue) in EZH2 KO relative to EZH2 Y726A on day 15.

GO terms associated with differentially expressed genes

By day 15 both EZH2 single mutants were found to upregulate genes associated with several metabolic and biosynthetic processes involved in generating nucleic acids and aromatic compounds, among others. In the EZH2 KO there was also an up regulation of genes associated with the processing of RNA as well as translation and gene expression. Both cell lines were found to downregulate many terms associated with neuronal development and cell signalling (Fig. 7.42 and 7.43).

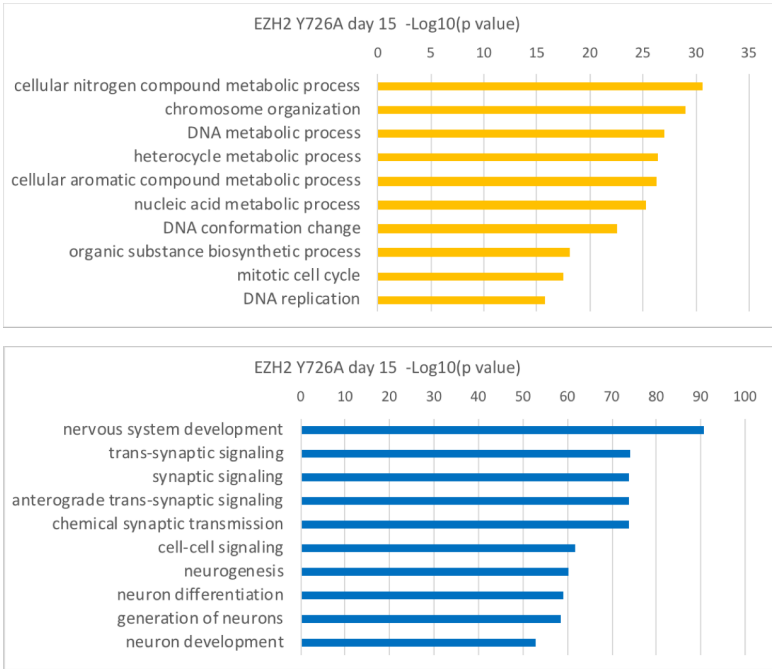


Figure 7.42 Gene ontology of genes differentially expressed on day 15 in EZH2 Y726A cells. Go terms enriched in genes significantly upregulated (yellow) and downregulated (blue) in day 15 EZH2 Y726A relative to day 15 E14 control.

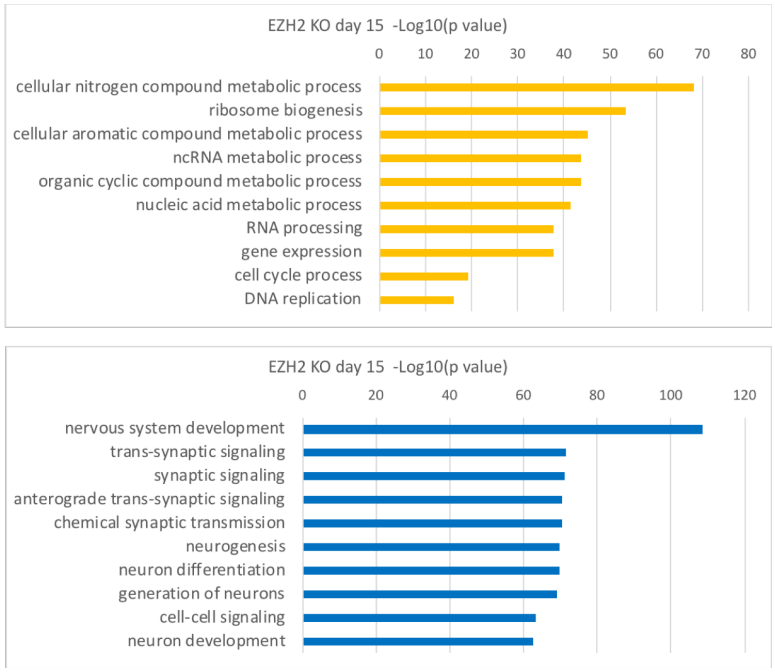


Figure 7.43 Gene ontology of genes differentially expressed on day 15 in EZH2 KO. Go terms enriched in genes significantly upregulated (yellow) and downregulated (blue) in day 15 EZH2 KO relative to day 15 E14 control.

Binding sites for different sets of TFs were enriched in each EZH2 single mutant amongst upregulated genes. The EZH2 Y726A was mostly enriched for binding sites of the ZF5 and E2F factors (Fig. 7.44). Surprisingly, some of the TFs whose binding sites were enriched amongst the upregulated genes in the EZH2 Y726A cells were also enriched in genes upregulated in the E14 control on day 15. In the EZH2 KO there was a larger variety of factors, including some Polycomb targets such as *Otx2* and *Pitx3*, as well as the ARNTlike factor, found to be downregulated on day 8 in the E14 control (Fig. 7.45). The downregulated genes in both EZH2 single mutants were enriched in binding sites for a similar set of TFs, which was also very close to those found in the genes downregulated on day 8, many of which are associated with neuronal development (Fig. 7.44 and 7.45). Of the genes found to be differentially expressed between EZH2 KO and EZH2 Y726A those that were upregulated in the EZH2 KO cells were found to be associated with response to LIF. Many genes found to be downregulated in the EZH2 KO cells as compared to EZH2 Y726A were associated with both cardiac and striated muscle (Fig. 7.41C). This suggests that the EZH2 KO may be behaving in a manner that bears some resemblance to a stem cell-like state. In contrast, the EZH2 Y726A cells although certainly divergent from the E14 control, may be behaving marginally more similarly to the differentiated state.

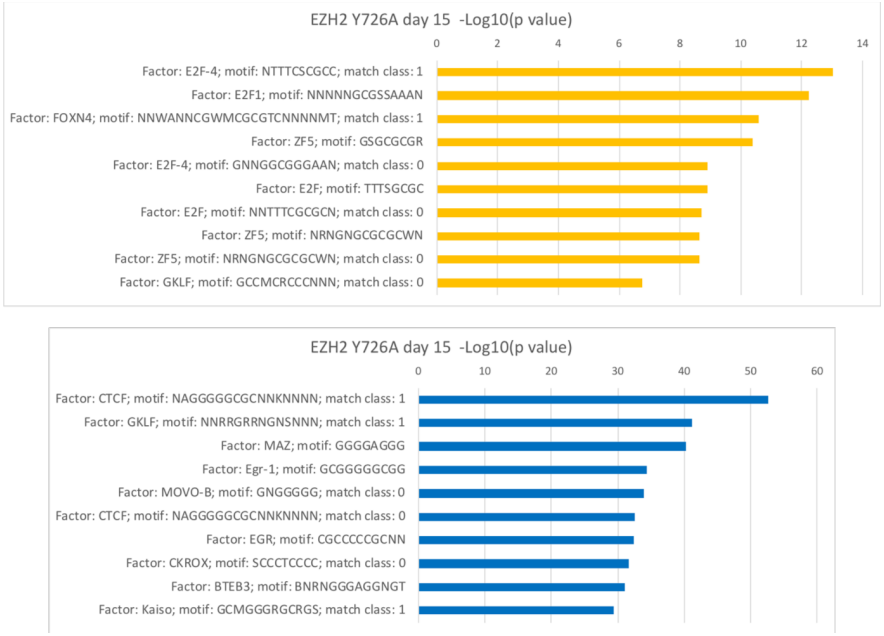


Figure 7.44 Enrichment of transcription factor binding sites in genes differentially expressed on day 15 in EZH2 Y726A cells. Transcription factor binding motifs enriched in genes significantly upregulated (yellow) and down regulated (blue) in day 15 EZH2 Y726A relative to day 15 E14 control.

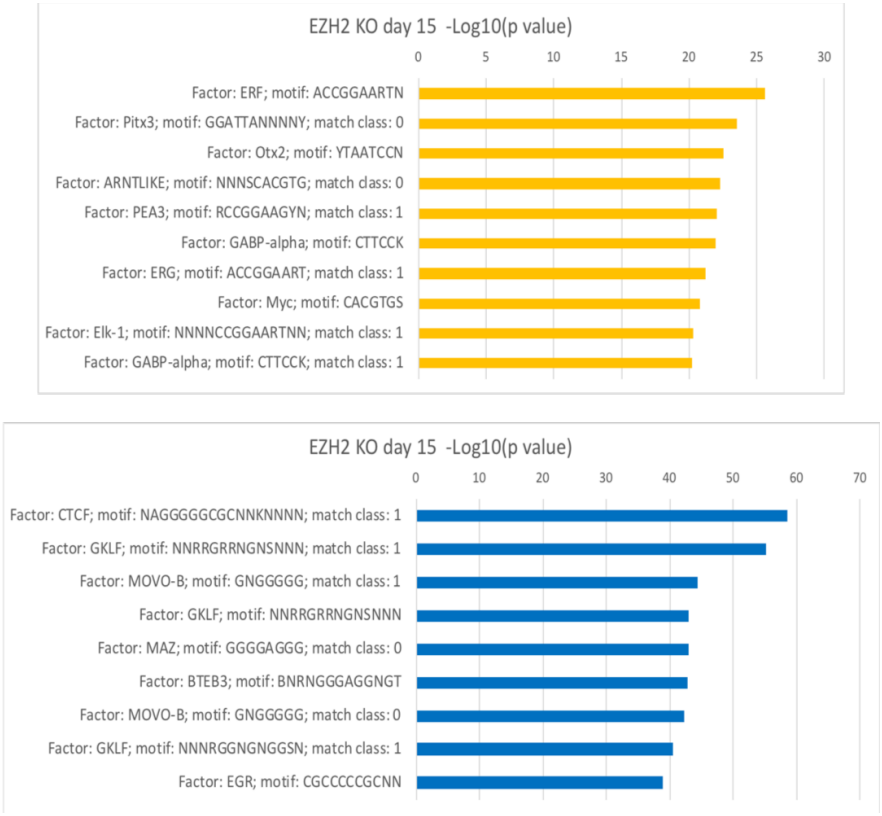


Figure 7.45 Enrichment of transcription factor binding sites in genes differentially expressed on day 15 in EZH2 KO cells. Transcription factor binding motifs enriched in genes significantly upregulated (yellow) and down regulated (blue) in day 15 EZH2 KO relative to day 15 E14 control.

Individual gene expression patterns

At the end of the differentiation protocol both the EZH2 Y726A and EZH2 KO cell lines were still overexpressing pluripotency factor *Oct4*, as was also seen by immuno-staining of day 15 cells (Fig. 7.8). EZH2 KO cells also overexpressed *Nanog* and *Sox2* on day 15 (Fig. 7.46). As expected at this timepoint, all genes used as markers both of neural progenitors and mature neurons were significantly underexpressed in both EZH2 single mutant cell lines, as is shown in Fig. 7.48 and 7.49. In the case of *Pax6*, although this gene was strongly downregulated in the E14 control at this stage, its expression was still significantly lower in both of the EZH2 single mutant cell lines on day 15. I also found that both *Otx2* and *Fgf5*, markers of early ectoderm development, were upregulated in both EZH2 single mutant cell lines on day 15 (Fig. 7.47), despite initially having been silenced on day 8. This may indicate a deficiency in the maintenance of a silenced state at these genes in the EZH2 Y726A and EZH2 KO.

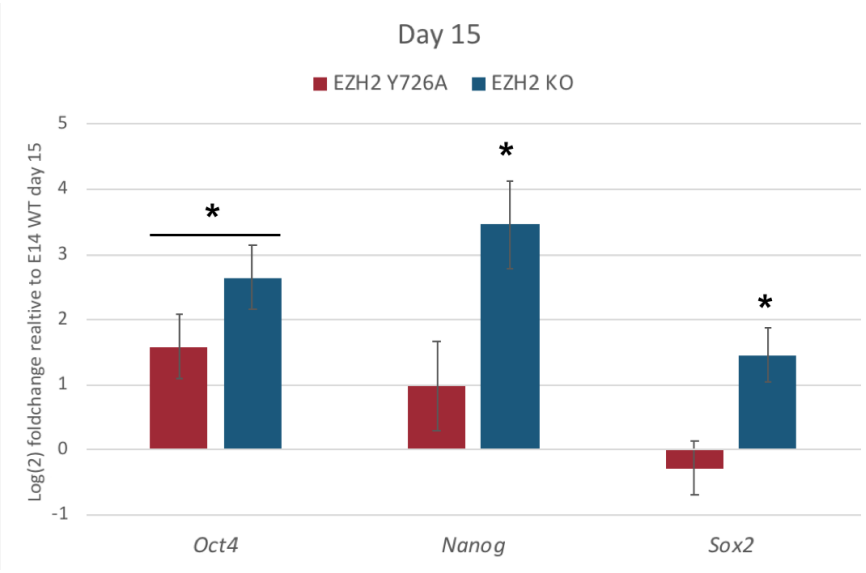


Figure 7.46 Expression of pluripotency factors in EZH2 singles mutant cell lines on day 15. Log(2) foldchange of expression of pluripotency factors in EZH2 Y726A and EZH2 KO cells on day 15 relative to day 15 E14 control. Significant changes in expression (p value $\leq$0.05) are marked by an asterisk.

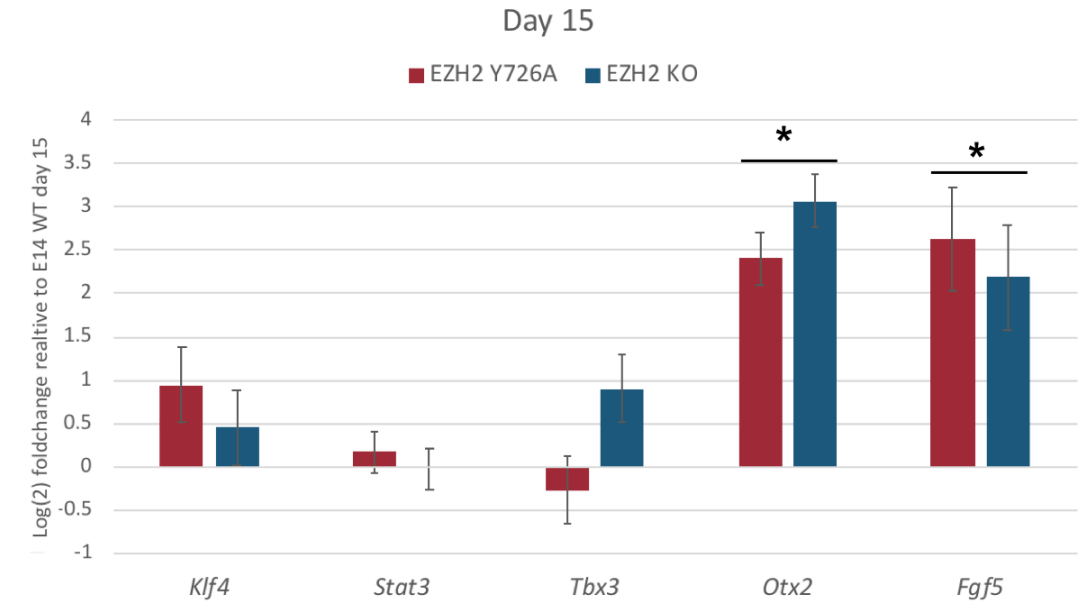


Figure 7.47 Expression of developmental genes in EZH2 single mutant cell lines on day 15. Log(2) foldchange of expression of genes known to be affected by withdrawal of LIF in EZH2 Y726A and EZH2 KO cells on day 15 relative to day 15 E14 control. Significant changes in expression (p value <0.05) are marked by an asterisk.

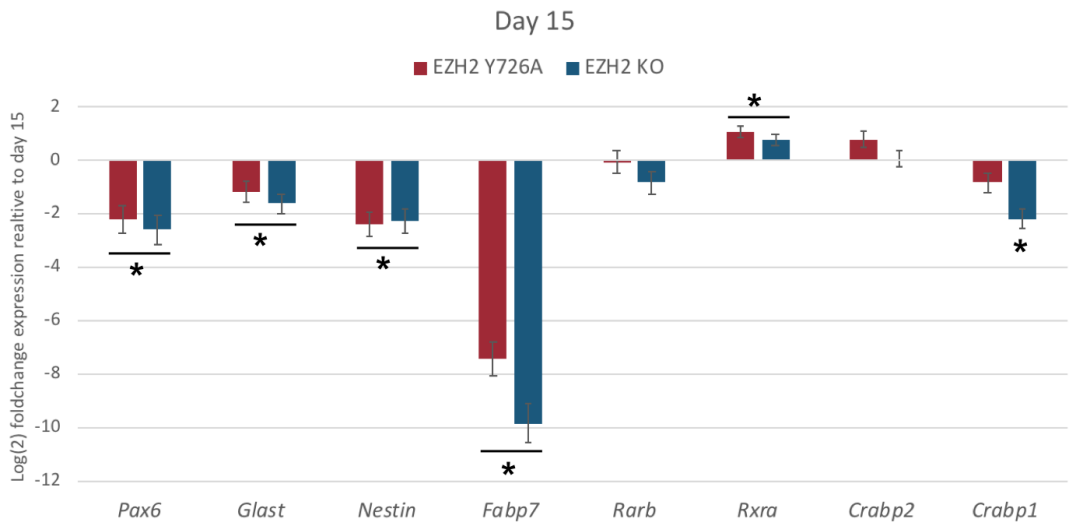


Figure 7.48 Expression of NPC markers and retinoic acid interactors in EZH2 single mutant cell lines on day 15. Log(2) foldchange of expression of neural progenitor marker genes and retinoic acid interacting genes in EZH2 Y726A and EZH2 KO cells on day 15 relative to day 15 E14 control. Significant changes in expression (p value $\leq$0.05) are marked by an asterisk.

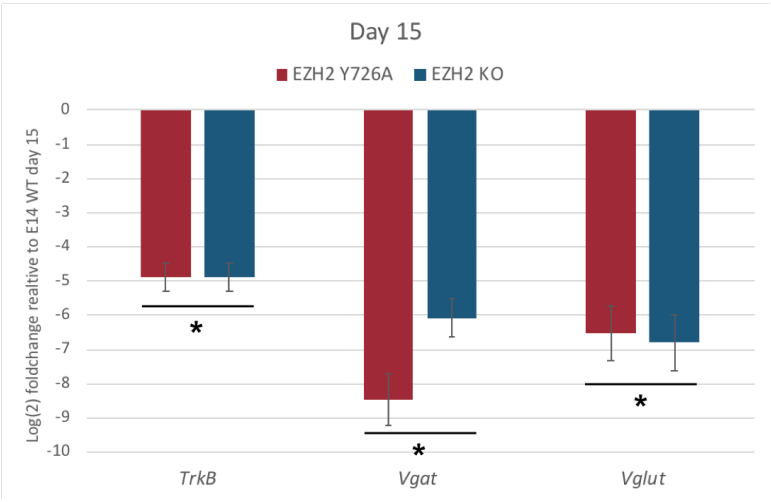


Figure 7.49 Expression of neuronal genes in EZH2 single mutant cell lines on day 15. Log(2) foldchange of expression of neuronal genes in EZH2 Y726A and EZH2 KO cells on day 15 relative to undifferentiated day 15 E14 control. Significant changes in expression (p value $\leq$0.05) are marked by an asterisk.

As previously discussed, the vast majority of EZH1/2 dKO cells did not survive until day 15 so the analysis of gene expression in these cells was not possible for this timepoint.

7.6 Conclusions

The experiments in this chapter demonstrated that all of the PRC2 mutant cell lines, both EZH2 single and EZH1/2 double mutants, failed to complete neural differentiation, although there were differences in viability of the single and double mutant cell lines, as well as in the genes misregulated. They further showed that the presence of a catalytically inactive EZH2 had little influence on gene expression as a whole throughout this process.

As expected, although all of the PRC2 mutants showed at least a slight difference in size and shape of EBs, the EZH1/2 dKO cell lines displayed a more exaggerated phenotype than the EZH2 single mutants. The latter only displayed marginally smaller EBs that the E14 control and were generally similar in shape while the former were significantly smaller in size and much more irregularly shaped, with higher rates of cell death. This was also the case for the EZH1/2 KO/Y726A cells, indicating that the catalytically inactive form of EZH2 was not able to compensate in any considerable way for the loss of EZH1. Indeed, upon dissociation of the EBs and plating of neural progenitor cells, there was a very low survival rate of the EZH1/2 double mutants, further suggesting that the catalytic activity of either EZH2 or EZH1 is critical for survival at this stage of differentiation. The EZH2 single mutant cell lines, in contrast, were able to survive and divide, although failed to form neuron-like cells. There was no discernible difference between the morphology of the EZH2 KO cells and the EZH2 Y726A cells at this stage, with both exhibiting a fibroblast like appearance devoid of neuronal protrusions, further suggesting that the catalytic activity of EZH2 is essential for neural differentiation.

Indeed, there was little difference in gene expression between either EZH2 single mutant by RT-qPCR analysis. Neither were able to sufficiently upregulate any of the neuronal markers assayed, and both failed to correctly repress pluripotency markers a trait which has been previously reported for EZH2 KO cells (Pasini et al., 2007; Shen et al., 2008). However, both mutants were capable of correctly expressing the early ectoderm marker *Otx2* in the early stages of differentiation, suggesting that the stage at which the EZH2 mutants become limiting for the progression of differentiation is after the addition of retinoic acid, whereas initial non-directed differentiation at the EB stage appears to be largely unperturbed.

The RNA-seq analysis provided further insight into the defects caused by the

mutations of EZH1 and EZH2. The principle component analysis showed that, in agreement with the RT-qPCR data, the biggest changes in the EZH2 single mutants occurred by day 8, further suggesting the addition of retinoic acid was a critical stage that requires a fully functioning PRC2 complex.

Consistently across every time point analysed, the EZH1/2 double mutants showed a higher misregulation of genes than either of the EZH2 single mutant cell lines. This indicated that although EZH1 is much less active than EZH2 and is not able to maintain levels of H3K27me3 seen in the E14 control, it is sufficient to maintain some level of gene regulation as opposed to when the PRC2 complex is lost completely. This could be down to the H3K27me3 that is still present in these cells, or may potentially arise from some other activity of the EZH1 protein itself.

It seemed that in general across all timepoints except day 15, there was very little significant difference between the gene expression profiles of EZH2 KO and EZH2 Y726A. Although the EZH2 KO cells were consistently found to have a larger number of differentially expressed genes as compared to the E14 control than the EZH2 Y726A, there were only ever a handful of genes found to differ significantly between the 2 mutants. This suggested that although the EZH2 Y726A mutant was still considerably divergent from the E14 control, the catalytically inactive complex may have been able to maintain a certain level of dampening to changes in expression at some genes. This effect was, however, subtle and much less significant than that conveyed by the presence of EZH1. One hypothesis for how this effect may be produced is through the simple obstruction of binding of activating factors by the presence of the catalytically inactive PRC2, a large protein complex, at promoters. It could also involve the interaction of PRC2 with additional factors that may themselves possess a repressive influence on transcription.

Despite EZH1 being able to compensate for EZH2 in mESCs and day 2 of differentiation, there were a few genes consistently differentially expressed in both of the EZH2 single mutant cell lines. Of these *Cdkn2a* has previously been shown to become upregulated in PRC2 depleted cells (Bracken et al., 2007). The majority of the remaining upregulated genes have CGIs in their promoter region and had at least a small enrichment of H3K27me3 in E14 control cells, although there was no common trait amongst these genes that I was able to identify that could explain why these select few were upregulated.

The comparison of gene expression in the EZH2 single mutant cell lines and the EZH1/2 dKO on day 8 showed that although neither were able to differentiate, the residual EZH1 in the EZH2 single mutant cells was able to maintain the repression of many Polycomb target genes, which was not observed in the EZH1/2 dKO. This suggests that the role of EZH2 at this stage of differentiation may not only involve silencing genes but also acting as an activator of certain factors, as seen by the large-scale downregulation (as opposed to the expected upregulation) of genes in the mutant backgrounds. This could either occur directly through the action of EZH2 in some capacity, or through the silencing of additional repressors to allow previously silenced genes to become expressed. This could also involve the pluripotency factors that were not entirely silenced in the PRC2 mutant cell lines. It has been shown that ectopic expression of these factors inhibits cell differentiation (Hochedlinger et al., 2005), so it may be that in these PRC2 mutants their overexpression was inhibiting the upregulation of certain factors required for commitment to the neural lineage. In both EZH2 single mutants I observed an upregulation of genes involved in spermatogenesis and meiosis on day 8. These genes have been shown to be enriched in H3K27me3 in mouse embryonic day 6.25 epiblast (Zylicz et al., 2015), which corresponds to the phase of embryonic development immediately prior to gastrulation. This could point to a failure of the EZH2 single mutants to maintain repression of this subset of Polycomb genes at this stage. Again, there appeared to be little difference between the EZH2 KO and EZH2 Y726A in how these genes were overexpressed.

Overall these differentiation experiments along with the RNA-seq data allowed me to gain some insight into when PRC2 is required during development. My data suggests that in the earlier stages of differentiation, the low levels of H3K27me3 produced by EZH1 are sufficient to maintain repression of Polycomb genes that would otherwise become derepressed. It also appeared that many of the changes in gene expression that happen in a wild-type background in the first few days of EB formation occur largely independently of PRC2. However, with the change in signalling cues from the addition of retinoic acid, the cells require a catalytically active EZH2 in order to correctly downregulate pluripotency factors and allow the expression of neuronal lineage genes. EZH1 may become insufficient at this stage because of its lower catalytic activity, or perhaps differences in how it is recruited, both or either of which may mean it is not able to bind to novel Polycomb target genes as cells differentiate. It could also be the case that there is a deficiency in the establishment of the required chromatin state earlier in differentiation, and these may only manifest as misregulated genes at a later stage.

Chapter 8

Conclusions and Discussion

Although there has been extensive research into the functions and mechanisms behind Polycomb group protein activity, the role of the H3K27me3 mark has not been completely distentangled from that of the PRC2 complex itself. The aim of this project was to establish a system in which these two aspects of PRC2 function could be decoupled in order to ascertain how much of Polycomb activity relies on the methylation of H3K27. The experiments described here showed that although the catalytic activity of EZH2 was largely dispensable to maintain the majority of Polycomb function in the presence of EZH1 in mESCs, it was essential for the correct regulation of gene expression upon differentiation of cells towards the neural lineage. In this chapter I will first discuss the findings on the state of Polycomb bound genes and the expression of the PRC2 complex itself in the PRC2 mutant cell lines, before detailing how this affects gene expression in the undifferentiated mESCs. Finally, I will discuss how these mutations in PRC2 affect the ability of these cells to differentiate into neurons, the consequences for gene expression, and the timing of PRC2 requirement throughout this process.

8.1 Characteristics of chromatin state at Polycomb target genes

While neither the catalytic inactivation of EZH2 nor a knockout of this protein appeared to have much of an effect on the expression of the other PRC2 subunits, the additional knockout of the EZH1 subunit did lead to some increased variation

in the expression of EED and SUZ12, with some clones expressing considerably lower levels of these subunits than the E14 control cells (see Fig. 3.10). A similar marginal decrease in expression of these subunits has recently been observed in independent strains of EZH1/2 dKO cells (Højfeldt et al., 2018; Lavarone et al., 2019) and decreased expression of other PRC2 subunits was noted in an earlier study with SUZ12 KO (Pasini et al., 2007). This may be the result of a feedback loop of the PRC2 complex in which decreased levels of the fully assembled complex may lead to decreased expression of the individual subunits. Alternatively this may simply reflect an increased stochasticity in gene expression in the EZH1/2 double mutants due to the lack of a major regulatory complex.

Despite these subtle effects on the expression of PRC2 subunits I found that the catalytically inactive EZH2 protein was still detectable at Polycomb target genes, both in the presence and absence of EZH1 and residual H3K27me3 (see Fig. 4.3). This indicates that PRC2 can be recruited to chromatin independently of H3K27me3, likely through a combination of the following mechanisms: the interaction with accessory factors, recruitment via the H2AK119ub mark placed by PRC1, and the binding of SUZ12 to CGIs (Hauri et al., 2016; Holoch and Margueron, 2017; Blackledge et al., 2014; Cooper et al., 2014; Højfeldt et al., 2018).

This binding of EZH2 Y726A to Polycomb target genes did however appear significantly reduced in the EZH1 KO background, and was still below wild-type levels even in the EZH2 Y726A single mutant cells. This suggests that, although not essential, H3K27me3 may contribute to the robust binding of PRC2 to its target sites, likely by providing an additional interaction site for the complex with chromatin, through the binding of the WD40 domain of EED (Margueron et al., 2009). There may also be an involvement of the EZH1 subunit itself in recruiting or stabilising the binding of EZH2 (Son et al., 2013). This may involve the dimerisation of PRC2 which can occur between two PRC2 complexes containing EZH1 and EZH2 respectively (Davidovich et al., 2014). A recent study, in which a catalytic mutant of EZH2 was generated by mutating the same residue targeted in this project, saw a similar maintenance of PRC2 binding to its target sites in the absence of H3K27me3 and EZH1 (Lavarone et al., 2019).

Once recruited to its target sites, PRC2 is thought to spread H3K27me3 across a broader region centred on the TSS of target genes via a mechanism which requires the allosteric activation of EZH2 which is triggered by the binding of EED to H3K27me3 (Margueron et al., 2009; Oksuz et al., 2018). The data

generated in the EZH2 ChIP-seq experiments should provide further information the requirement of this mechanism for the binding of PRC2 to its target sites, specifically by comparing the spread of the EZH2 enrichments across target genes in EZH2Y726A cells to those seen in the E14 control cells. If EED-bound H3K27me3 is required for spreading of PRC2 activity from its initial recruitment site the EZH2 signal should be broader in the E14 control cells than in the EZH2 Y726A. Additionally, Oksuz et al. classified Polycomb targets into nucleation sites, which they defined as the primary target sites for Polycomb binding, and spreading sites that required physical contact with these nucleation sites to acquire PRC2 binding and H3K27me3 deposition (Oksuz et al., 2018). This spreading of PRC2 from the nucleation sites was found to be lost in cells with a mutation in the EED subunit that abolished its binding of H3K27me3. It would be interesting to determine whether EZH2 Y726A is confined to the nucleation sites identified in this study, or if it is also found at the so called spreading sites, which would suggest that there are additional factors enabling its recruitment to these targets.

Future experiments that will be undertaken concerning PRC2 binding to chromatin will involve cell lines recently generated in the lab in which we have introduced a mutation to disable the catalytic activity of the EZH1 subunit. The comparison of EZH2 binding to it targets in EZH1/2 KO/Y726A cells and in cells with a catalytically inactive EZH1 in a EZH2 Y726A background will help further elucidate the role of EZH1 in EZH2 binding to chromatin independently from H3K27me3.

As has been previously described (Margueron et al., 2008; Lavarone et al., 2019), I found that very little H3K27me3 was deposited by EZH1 in the absence of EZH2 (see Fig. 4.2). Further analysis of the H3K27me3 ChIP-seq data may provide more insight into which genes retain the most and the least H3K27me3 in absence of EZH2, and whether this correlates at all with the genes found to be misregulated in these cells. This experiment will reveal more information about the target preference of EZH1 as compared to EZH2, and the comparison of the EZH2 KO to the EZH2 Y726A cells should demonstrate whether the presence of the catalytically inactive EZH2 at these sites has any influence on the activity or the targeting of EZH1. It will additionally be interesting to compare the sites that retain the most H3K27me3 in the EZH2 single mutant cells lines to the so called nucleation sites defined in the study by Oksuz et al. As EZH1 is known not to respond to the allosteric activation provided by EED in presence of H3K27me3,

it might be expected that it would only be found at the nucleation sites and not the spreading sites defined in this study (Oksuz et al., 2018; Lee et al., 2018).

A surprising effect of the EZH2 single and EZH1/2 double mutations was that on the binding of RING1B. As was shown in Fig. 4.5, in both EZH2 single mutant cell lines there was a slight reduction of RING1B binding to the set of Polycomb genes assayed, however this effect was much more striking in the EZH1/2 double mutant cells, with little to no RING1B bound in these cell lines. This suggests that the recruitment of PRC1 to these genes occurs predominantly via the hierarchical model, in which the H3K27me3 placed by PRC2 allows the binding of PRC1 through recognition of this mark by the CBX subunit (Wang et al., 2004). The small reduction in binding of RING1B observed for the EZH2 single mutants was surprising given how little H3K27me3 remains at these sites (see Fig. 4.2), suggesting that very low levels of H3K27me3 are sufficient to maintain recruitment of canonical PRC1. This finding indicates that Polycomb silencing is a mechanism that is able to withstand small variations in expression of the subunits or in catalytic activity without this strongly affecting the levels of expression of its target genes. This is an unsurprising characteristic for a regulatory network that is so crucial in development, particularly as many of the genes targeted by Polycomb, such as the Hox genes, are themselves developmental regulators, small changes in the expression of which can have considerable effects on tissue specification and body patterning in the developing embryo (Lewis, 1978; Zink and Paro, 1989; O'Carroll et al., 2001; Pasini et al., 2004). This finding, in addition to the results from chapters 5 and 6, also helps to explain why EZH1 is able to fully take over from EZH2 once cells are terminally differentiated despite its much lower catalytic activity, by showing that even in cells where EZH2 is the dominant subunit, EZH1 is able to maintain the vast majority of Polycomb function in cells that are not transitioning through differentiation (Laible et al., 1997; Bracken, 2003; Margueron et al., 2008).
It would be interesting to assess the binding of RING1B genome wide to determine if this effect is seen at all Polycomb sites, particularly those for which there is no H3K27me3 detected in the EZH2 single mutant cells. The double EZH1/2 catalytic mutant cells may also be of use here as a further confirmation that the low levels of H3K27me3 placed by EZH1 are indeed sufficient to recruit PRC1.

A recent study in which a conditional catalytic mutant of PRC1 was generated found that the catalytic activity of this complex was essential for the majority of Polycomb gene repression as well as Polycomb domain formation and long-range

interactions between Polycomb target sites (Blackledge et al., 2019). Additionally this study also found that the loss of PRC1 catalytic activity significantly decreased PRC2 and H3K27me3 occupancy at target sites. While the binding of the canonical PRC1 complex to target sites was found to be dependent on PRC1 catalytic activity, the occupancy of the variant form of PRC1 was found to be largely unaffected by the loss of catalysis. These results reaffirmed the model in which the binding of PRC1 and PRC2 is mutually reinforced at target sites where a low level of transcription enables significant deposition of H2AK119ub, likely first initiated by the variant form of PRC1. This study is in agreement with my findings regarding the requirement for PRC2 activity for the recruitment of PRC1 and the role of PRC1 in mediating the *Lhx5* enhancer-promoter contact. In light of the results of this study it would be interesting to assess the levels of H2AK119ub at Polycomb genes, particularly in the EZH1/2 double mutants in which I found the RING1B binding to be reduced to undetectable levels while maintaining binding of the EZH2 Y726A mutant (see Fig. 4.3 and 4.5). This experiment would provide insight into the contribution of H2AK119ub to the recruitment of PRC2 to these sites, and how the loss of EZH2 catalytic activity may affect the ability of PRC1 to deposit its mark.

Additional changes to histone modifications were noted in the PRC2 mutants, notably for H3K4me3 which co-localises with H3K27me3 at bivalent genes, and which was studied in more detail in chapter 4. The H3K4me3 mark was found to increase at bivalent genes in all PRC2 mutant mESCs, although this increase was the most striking in the EZH1/2 double mutant cells (see Fig. 4.4). As H3K4me3 and H3K27me3 are known to be deposited by mutually antagonistic processes (Schmitges et al., 2011; Voigt et al., 2012), this result was to be expected given the decrease of H3K27me3 that occurs in each PRC2 mutant cell line, with the most complete depletion of this mark induced in the EZH1/2 double mutant cells. One surprising finding was that the increase in H3K4me3 at bivalent genes appeared to be stronger in the EZH1/2 KO/Y726A cells than in the EZH1/2 dKO. This result suggests that the catalytically inactive PRC2 bound to these genes may somehow favour the deposition of H3K4me3. This correlates with some unpublished *in vitro* experiments that have been performed in the lab in which the presence of a catalytically inactive PRC2 complex led to an increased activity of MLL2 in catalysing H3K4me3. If this phenomenon is confirmed by further analysis of the H3K4me3 ChIP-seq data it could contribute to the explanation of how a bivalent chromatin state is established, or how H3K4me3 is maintained in ESCs at sites which carry inherently repressive features.

Alternatively this increase in H3K4me3 may reflect the activation of transcription of these genes, which may involve multiple H3K4 methyltransferases. In order to confirm the role of MLL2 in this phenomenon, inactivation of this complex, by knockout, mutation of the SET domain or an inhibitor will be required in addition to the PRC2 mutations to verify that H3K4me3 is indeed deposited by this complex alone at these sites.

A recent study found a similar positive effect on methylation from the presence of a catalytically inactive methyltransferase, Dnmt3b (Nowialis et al., 2019). This study compared a catalytically inactive Dnmt3b to a Dnmt3b KO in mESCs and found that this protein had accessory functions that were independent of its catalytic activity which largely rescued the phenotype of the KO mutant. Catalytically inactive Dnmt3b was found to be able to restore the majority of the DNA methylation lost upon the full KO of this protein, possibly through a non-catalytic function which supports the methyltransferase activity of other Dnmts. Unlike my findings for EZH2 Y726A however, the catalytically inactive Dnmt3b maintained wild-type expression levels for the majority of genes which underwent misregulation in Dnmt3b KO cells. This was found in many cases to be due to the maintenance of DNA methylation by the catalytically inactive Dnmt3b, but for some genes, such as *Wnt9b*, due to a direct repressive effect of Dnmt3b binding to gene promoters independent of DNA methylation. As previously stated, the analysis of the ChIP-seq data for H3K27me3 in the EZH2 single mutant cells may reveal whether EZH2 Y726A has a comparable accessory function in directing the methyltransferase activity of EZH1 to target sites.
Further analysis of gene expression differences between the EZH2 Y726A and EZH2 KO cells at sites which are EZH2 positive but H3K27me3 negative in the E14 WT cells may also reveal any direct inhibitory effects of EZH2 that would be entirely independent of catalysis, analogous to the effect of Dnmt3b on *Wnt9b* observed in this study.

The experiments in chapter 5 demonstrated that, similarly to all other Polycomb targets assayed previously in this project, PRC1 is primarily recruited to the *Lhx5* gene by its binding to H3K27me3 (see Fig. 5.2B). These experiments also showed that although the contact between the *Lhx5* enhancer and promoter was lost in the absence of the PRC2 complex (see EED KO and EZH1/2 dKO Fig. 5.5), this contact was maintained in the presence of EZH1. However, this contact was additionally shown not to be directly mediated by PRC2 but PRC1 as the knockout of RING1B led to dissociation of the *Lhx5* enhancer and promoter,

but did not result in a reduction of H3K27me3 levels at these sites (see Fig. 5.5 and 5.2A). The additional loss of contact observed in the CBX7 F11A mutant further confirmed that at this site, PRC2 is required only insofar as it enables the recruitment of PRC1. Interestingly, although this contact has been shown to be required for the timely expression of this gene upon cell differentiation (Cruz-Molina et al., 2017), it would appear from these experiments that the dissociation of this contact upon loss of Polycomb also leads to upregulation (see Fig. 5.6). This Polycomb dependent contact of the enhancer and promoter of *Lhx5* may then serve the dual purpose of both maintaining transcriptional repression of this gene in ESCs, whilst also setting up a conformation that will enable its expression at the correct moment upon differentiation. The results in this chapter are preliminary and require additional biological replicates as well as the inclusion of the EZH1/2 KO/Y726A cells. Again, the EZH1/2 double catalytic mutant may be of use in these experiments to rule out any role of the EZH1 protein itself in these contacts besides the H3K27me3 that it catalyses. As this was only performed for the enhancer-promoter contact of one gene it cannot be ruled out that PRC2 may be directly involved in the contact of different types of long-range contacts. In the longer term, it would be interesting to expand this study to additional sites that have been found to form PRC2 clusters and/or long-range contacts with other sites, or even perform a genome wide study using one of the "C" methods of chromatin conformation capture. This may be of particular interest for the subset of Polycomb targets that are not bound by the PRC1 complex (Ku et al., 2008).
Additionally, further study of Polycomb-dependent contacts at different stages of differentiation may be interesting. As EZH1 becomes insufficient to maintain proper gene expression throughout this process it may also lose its ability to maintain contacts in this context as well.

8.2 Consequences of EZH2 and EZH1 mutations on gene expression and pluripotency in mESCS

Following the evaluation of the chromatin state at Polycomb bound genes in the PRC2 mutant cell lines, I next moved on to assess the effects these mutations had on gene expression. In undifferentiated mESCs, as has been previously reported

(Pasini et al., 2007; Chamberlain et al., 2008; Shen et al., 2008; Riising et al., 2014), the loss of EZH2 alone had a limited effect on gene expression and it was only when the activity of both EZH1 and EZH2 were lost that large-scale changes in gene expression occurred (see Fig. 6.3, 6.4 and 6.5). This appeared not to affect the expression of pluripotency factor *Oct4* (see Fig. 6.1) and self renewal of these cell lines appeared to be largely unaffected as well. These results along with those of previous chapters, further demonstrate that in mESCs, EZH1 is largely able to compensate for EZH2 in the maintenance of a Polycomb silenced state, despite its much lower catalytic activity. It remains to be seen however, whether EZH1 would also be able to rescue all PRC2 function if it were re-expressed in the EZH1/2 dKO cells. It may be that the maintenance of Polycomb gene silencing achieved by EZH1 in the EZH2 single mutant cell lines, may only have been possible because these sites were already established as Polycomb bound genes.

In both the presence and absence of EZH1, the presence of EZH2 Y726A produced a very similar gene expression pattern in cells as the EZH2 KO, demonstrating that the presence or absence of H3K27me3 is much more influential on Polycomb gene regulation than the presence or absence of the complex itself. The RNA-seq data did seem to suggest that there might be a higher degree of variance in gene expression from sample to sample in the knockout background as compared to the EZH2 Y726A background, but as there were only 3 biological replicates for each genotype it was difficult to assign any statistical significance to this finding. There was however, both in the EZH2 single mutants and the EZH1/2 double mutants, a larger number of genes misregulated in the EZH2 KO background than the EZH2 Y726A background. This suggests that the catalytically inactive EZH2 may have a small effect on gene expression by its continued binding at these sites. This could occur simply by its presence on chromatin blocking the binding of additional transcription factors, or perhaps by inhibiting the progression of Pol II along the gene. As PRC2 has been shown to interact with nascent RNAs this may also be happening with EZH2 Y726A, which could disrupt the processing and maturation of these transcripts (Kaneko et al., 2013; Kaneko et al., 2014; Skourti-Stathaki et al., 2019). Alternatively the presence of PRC2 at these genes may maintain a certain level of clustering of Polycomb genes together into a less transcriptionally active chromatin domain.

There were a small number of genes that were misregulated in the EZH2 single mutants, most of which were affected in both the EZH2 KO and EZH2 Y726A. It would be interesting to investigate further why these genes in particular are

affected by the loss of EZH2 activity alone. It might also be interesting to see how the state of Pol II occupancy is affected in these cell lines as it is known that is is bound in a poised state at the promoters of bivalent genes (Stock et al., 2007; Brookes et al., 2012).

8.3 Consequences of EZH2 and EZH1 mutations on neural differentiation

The aim of the neural differentiation experiments in chapter 7 was to clarify the requirements of PRC2 during this process. It was evident from these experiments that H3K27me3 is essential for neural differentiation, specifically that deposited by EZH2. This is in contrast to the previous results in this project which demonstrated that the residual levels of H3K27me3 placed by EZH1 are largely sufficient to maintain Polycomb repression in mESCs.
The results of these experiments additionally suggested that PRC2 is largely dispensable for many of the events that occur within the first few days of embryoid body formation, as the number of genes significantly misregulated in the PRC2 mutants as compared to the E14 control did not increase between day 0 and day 2 of the differentiation (see Fig. 7.21). Instead, PRC2 seemed critical for the changes in gene expression that occur in response to the addition of retinoic acid (see Fig. 7.26). A critical role of PRC2 at this point during the differentiation is perhaps somewhat surprising given this result that both EED and SUZ12 see a decrease in their expression levels at this same timepoint in the E14 control (see Fig. 7.19), suggesting a downregulation of PRC2 activity around that stage of the protocol. This later timing for PRC2 requirement is perhaps surprising given the early lethality of Polycomb knockouts in embryonic development (Faust et al., 1998; O'Carroll et al., 2001; Pasini et al., 2004). However, the experimental conditions of this differentiation protocol are considerably different to the environment in the early embryo so the comparison between the two situations is not easy to make.

Many of the genes that were differentially expressed during differentiation in the PRC2 mutant cells were developmental genes from the neural lineage. This was not necessarily the expected phenotype, as PRC2 is a repressive complex, compromised silencing would seem more likely than an inability to upregulate genes. This further suggests that the bivalent state at many of these genes is

required for their induction at the correct time during differentiation.

From both the gene expression patterns and cell morphology it was clear that the EZH1/2 double mutant cell lines had a greater, and earlier, defect in their ability to differentiate than the EZH2 single mutant cells (see Fig. 7.2, 7.3 and 7.4). It was found that while the EZH1/2 dKO mutation led to the upregulation of many Polycomb target genes both in ESCs and throughout differentiation, this was not the case in the EZH2 single mutants. This demonstrated that while EZH1 was not able to compensate completely for EZH2 throughout this process, it was able to maintain a certain level of repression at genes that were Polycomb bound in the undifferentiated ESC state. An example of this can be seen in Fig. 7.39: the Hox genes that do not become actively transcribed in the E14 control cells are not derepressed in either of the EZH2 single mutant cell lines, however there is an overexpression of these genes in the EZH1/2 dKO. For those Hox genes that were strongly upregulated in the E14 control on day 8 however, neither the EZH2 single mutant cells nor the EZH1/2 dKO expressed these genes to the level seen in the E14 control, despite all of these genes being upregulated in the EZH1/2 dKO as compared to the E14 control in the undifferentiated state (see Fig. 6.12). This suggests that even if the Polycomb repression is removed from these genes, they require an additional activating factor to achieve the levels of expression seen in the E14 control on day 8. This factor appears to be missing in both the EZH2 single and the EZH1/2 dKO cells. This could be because a repressor of this activator is not being silenced in these cells. In this case one of the ways in which EZH1 is not able to compensate for EZH2 during differentiation may be an inability to locate to new targets or at least to establish a Polycomb repressive state at these loci where repression is newly required.

One theory to explain this deficiency involves EZH1's much reduced response to allosteric activation by EED (Lee et al., 2018). This effect has been shown to be reproduced by the binding of the methylated K116 of JARID2 by EED (Sanulli et al., 2015; Kasinath et al., 2018) which has been proposed to aid in the targeting of PRC2 to novel, previously unmethylated sites. In the case of EZH1 this stimulation of activity would not occur, perhaps impairing its binding to new sites. However, there are additional methods of recruitment for PRC2 that do not require H3K27me3, for example it has been recently shown that the N-terminal VEFS domain of SUZ12 is able to bind CGIs in the absence of H3K27me3 (Højfeldt et al., 2018). The existence of these additional modes of recruitment suggest the deficiencies of EZH1 during differentiation may not be due to compromised targeting of EZH1 but rather due to its reduced activity

being insufficient for the robust establishment of an enrichment of H3K27me3 at these sites.

Another factor possibly preventing the upregulation of neuronal lineage genes in the PRC2 mutant cells may be the higher expression levels of pluripotency factors such as *Oct4* and *Nanog* (see Fig. 7.34). Their continued expression may be maintaining the activation of repressors that inhibit the expression of factors involved in the response to retinoic acid and neural differentiation in general. However, though the expression of these pluripotency factors was found to be significantly higher in the PRC2 mutant cells than in the E14 control by day 8 (around 5 times higher in the EZH2 single mutant cells and 10 times in the EZH1/2 dKO), these levels were still much reduced as compared to those detected in the undifferentiated mESCs, approximately 10 times lower for *Oct4* and around 5 times lower for *Nanog* (Fig. 7.35 and 7.36). This has been previously observed in EED KO cells upon embryoid body formation (Obier et al., 2015). This study also showed that many of the genes that acquire H3K27me3 during EB formation were bound by OCT4 in ESCs, further demonstrating the antagonistic activity of this factor to Polycomb regulation of gene expression. Given these findings, it is conceivable that the low amount of pluripotency factors remaining in the PRC2 mutant cells by day 8 may be sufficient to maintain some expression of their target genes, particularly if said genes are not being efficiently bound by PcG proteins.
To further investigate this, ChIP experiments could be performed at this time point to determine if EZH1 is able to bind to novel targets and deposit H3K27me3. ChIP for OCT4 and NANOG could also be performed to ascertain whether the low levels of expression detected for these genes by RNA-seq in the PRC2 mutant cells translate to detectable levels of these proteins still bound to their target genes by day 8. Although these factors are known to be bound by Polycomb in differentiated cells (Pasini et al., 2007; Obier et al., 2015), their downregulation in the EZH1/2 dKO cells demonstrates that their downregulation is mediated by factors independent of PRC2, suggesting that Polycomb functions here as a marker of already silenced genes, serving to maintain the repressive state rather than establishing it, as has been previously demonstrated (Riising et al., 2014). A similar observation was made in early studies of PcG proteins in *Drosophila* in which they were found to be dispensable for the establishment of a repressed state and only required for its maintenance (Struhl and Akam, 1985; Jones and Gelbart, 1990; Simon et al., 1992).

As was observed in the undifferentiated mESCs, the EZH2 Y726A cells misregulated slightly fewer genes than the EZH2 KO at each timepoint studied (see Fig. 6.4, 7.21, 7.26 and 7.40). This further highlights the possibility that EZH2 Y726A may be able to maintain a certain level of regulation of gene expression that is lost in the EZH2 KO and is independent of H3K27me3. As was discussed above, there are several potential mechanisms for how this might be achieved by the catalytically inactive complex. Further investigations into how this process occurs could be done with a reporter gene to which PRC2 could be artificially recruited. Expression levels of this gene could then be compared in presence of a wild-type or catalytically inactive PRC2 or a PRC2 KO. By adding transcription factor binding sites to this reporter gene, we could also determine whether a catalytically inactive PRC2 can significantly affect the recruitment of transcription factors. This assay could also be used to determine whether binding of a catalytically inactive PRC2 is sufficient to co-localise the reporter gene with other Polycomb targets within the nucleus.

To further investigate the requirements of PRC2 throughout the different stages of differentiation we intend to make use of the auxin-inducible degron system. This system evolved in plants and leads to the rapid degradation of proteins by ubiquitylation in response to auxin, a plant hormone, and has been developed to function in multiple different cell types (Nishimura et al., 2009). We have generated cell lines with an auxin degron-tagged EZH2 which enables us to induce the degradation of this protein upon treatment with auxin. Using these cell lines neural differentiation experiments will be performed throughout which EZH2 will be removed at different stages. The aim of these experiments will be to determine whether PRC2 truly is dispensable in the early stages of EB formation before the addition of retinoic acid, as suggested by the differentiation data, or whether it is perhaps needed to establish the correct chromatin state at this time that is required for proper gene expression later. These experiments will also determine whether PRC2 is required once cells have already established a *Pax6* expressing neural progenitor state, for the continued differentiation towards mature neurons. It is conceivable that once certain key genes have been expressed, regulatory networks are activated that will control the expression of the required genes for the continued progression towards the neural lineage without any major requirements from PRC2.

As the loss of H3K27me3 induced such strong changes in gene regulation, it proved difficult to discern the more subtle effects that could be mediated by the

PRC2 complex itself. One approach to further explore the potential catalysis-independent functions of the complex would be to use a system in which PRC2 is absent from its target genes but the H3K27me3 is maintained. We hope to use the auxin degron-tagged EZH2 cells for these experiments, as the protein is degraded within the first 24 hours of treatment, but H3K27me3 is maintained for at least 48 hours, giving us a window in which genes still retain the histone modification but are now devoid of PRC2.

An interesting additional PRC2 mutation to study would be an overactive EZH1. Some cancers have been found to harbour overactive EZH2 mutants that with higher catalytic activity than the wild-type protein (Morin et al., 2010; Bodor et al., 2013; Harms et al., 2014). This has also been identified in one case for EZH1 (Calebiro et al., 2016). Introducing this mutation into an EZH2 KO or EZH2 Y726A background would be interesting from several perspectives. First, this could help establish whether there are any substantial differences in targeting or additional activities of EZH1 and EZH2. Second, by using these cells in differentiation assays, it would be possible to determine whether the lower catalytic activity of wild-type EZH1 is the only reason it is not able to compensate for EZH2 during this process or whether there are additional factors at play. Additionally, this overactive form of EZH1 could be expressed in EZH1/2 double mutant cells to test whether it is sufficient to re-establish Polycomb repression in these cells.

8.4 Summary and final remarks

Overall this project provided a greater understanding of how the PRC2 complex functions, both in mESCs and throughout neural differentiation. Although it began with a focus on the EZH2 subunit, the early findings of this thesis broadened the scope of the project to include questions regarding the separate functions of EZH2 and EZH1, as the latter proved to be of more consequence in undifferentiated cells than originally anticipated. I found that in mESCs very little H3K27me3 is required for the maintenance of Polycomb function, including gene repression, the recruitment of PRC1 and the contact of the enhancer and promoter of Polycomb target gene *Lhx5*, and that EZH1 alone was capable of fulfilling this requirement. However, it remains possible that EZH1 has further roles beyond placing residual H3K27me3 that can contribute to maintenance of repression. In the absence of any H3K27me3, I found that Polycomb genes became

derepressed in mESCs, and this was only weakly mitigated by the presence of a catalytically inactive EZH2. This indicates that the catalytic activity of PRC2 is essential for the vast majority of its function in mESCs.

In neural differentiation I found that in the earliest stages of EB formation, the effect of the loss of PRC2 function on gene expression did not differ greatly from that observed in the undifferentiated mESCs. The presence of EZH1 and the residual H3K27me3 was sufficient to maintain the correct expression of the vast majority of genes at this early timepoint. However, the catalytic activity of EZH2 was essential for correct gene expression upon treatment with retinoic acid. During this process EZH1 was able to maintain repression of genes that were Polycomb targets in mESCs, however it did not allow for the upregulation of genes of the neural lineage, nor full repression of pluripotency factors. The presence of the catalytically inactive EZH2, again, was not able to maintain correct gene expression patterns required for neural differentiation, but there were a consistently lower number of genes differentially expressed in these cells than in the EZH2 KO cells, suggesting that PRC2 containing catalytically inactive EZH2 is able to maintain binding and some degree of function.

Based on the findings of this thesis I propose the following working model: in undifferentiated cells in which the set of genes targeted by the Polycomb group proteins does not change, low levels of H3K27me3 are sufficient to maintain the required repression and chromatin state at Polycomb target genes, and EZH1 is sufficient for this purpose. However, in a dynamic system, such as during differentiation, where Polycomb binding is required at novel sites, EZH2 is required for the binding of PRC2 to new genes and depositing high levels of H3K27me3, allowing for the robust establishment of Polycomb repression at these sites.

Chapter 9

Appendix

ChIP antibodies	Species	Source	Identifier
H3K27me3 (C36B11)	Rabbit monoclonal	Cell Signaling Technology	9733
H3K4me3 (C42D8)	Rabbit monoclonal	Cell Signaling Technology	9751
EZH2 (D2C9)	Rabbit monoclonal	Cell Signaling Technology	5246
RING1B (D22F2)	Rabbit monoclonal	Cell Signaling Technology	5694
WB antibodies			
LAMIN A/C (H-110)	Rabbit polyclonal	Santa Cruz	SC20681
EZH1	Rabbit polyclonal	Margueron et al. 2008	
EZH2 (11)	Mouse monoclonal	BD Biosciences	612666
H3K27me3 (C36B11)	Rabbit monoclonal	Cell Signaling Technology	9733
H3K4me3 (C42D8)	Rabbit monoclonal	Cell Signaling Technology	9751
H3K27ac (D5E4)	Rabbit monoclonal	Cell Signaling Technology	8173
SUZ12 (D39F6)	Rabbit monoclonal	Cell Signaling Technology	3737
RING1B (D22F2)	Rabbit monoclonal	Cell Signaling Technology	5694
CBX7	Rabbit polyclonal	Abcam	AB21873
OCT4	Rabbit polyclonal	Sigma-Aldrich	P0056
H3	Rabbit polyclonal	Abcam	AB1791

Figure 9.1 Table of all antibodies used in this thesis

EZH2 Y726A	
EZH2 Y726A ssDNA HR donor	CTGCTCACTGCTGCCAATGTTTCTCTTTGCAGTTATGATGGTTAATGGTGACCACAGGATAGGCATCTTTGCTAAGAGGGCCATA CAAACCGGTGAGGAGTTGTTTTTTGATGCCAGGTTGGTAAAATACACAAAGGTTTTCCCTTCTGAAACTATGGGGCTTGGGTCA AAGGCTGTTTGTAAGTTTTTTGACAG
guide RNA sequence	GCTAAGAGGGCTATCCAGAC

EZH2 KO ex6	
EZH2 KO ex6 ssDNA HR donor	AAATCAGCTTTGTTGAGATGCTGTGATACAGAGAAAATGAAATTGAAACTTGTTTCTTCCATCTCTTCCCTTATCACAATTCACT GTTTAGATTAATAAGCTTGATCACCTCGGAAATTTCCTGCTGATAAAATATTTGAAGCCATTTCCTCAATGTTTCCAGATAAGGG CACCGCAGAAGAACTGAAAGAAAAGTAAG
guide RNA sequence	CAGCAGGAAATTTCCGAGGT

EZH1 KO ex8	
EZH1 KO ex8 ssDNA HR donor	TGTTCTGCTGCAACAAAGGGAAAACGACAAACAGCTTTGTCTATGCCCAGGAGTTCACTGTCATTTCCCATTCAGGCAACTAAT AGGGATCCTAGTAACAGTTTCCAAACCATATGATATTCAGCGCCATTGCGTCCATGTTTCCTGAGAATGGTGTCCCTGACGACA TGAAGGAGAGGTAGGGACAGCTGTTCCACTCA
guide RNA sequence	GCGCTGAAGATCATGTCATT

CBX7 F11A	
CBX7 F11A ssDNA HR donor	AGCTCCACCCCCGGCCCAGCCCGGTCCGGTCCTGGGCCCACTCGCCCGCCGCCCCGCATGGAGCTGTCAGCCATAGGCGAACA AGTGGCTGCGGTGGAGAGCATCCGGAAGAAGCGCGTGCGGAAGGTGAGGCTGCTCGGGGGCGGCTCCCGGGACAGGGCGAG GTTGCCTCCCGCTCC
guide RNA sequence	GCTGTCAGCCATAGGCGAGC

Figure 9.2 Table of CRISPR gRNA and dsDNA homologous repair templates

Gapdh ChIP F	GGGTTCCTATAAATACGGACTGC
Gapdh ChIP R	CTGGCACTGCACAAGAAGATG
Oct4 ChIP F	GGCTCTCCAGAGGATGGCTGAG
Oct4 ChIP R	TCGGATGCCCCATCGCA
Hoxc5 ChIP F	GTACTGCTACGGCGGATTGG
Hoxc5 ChIP R	TACCCCGTGGAGAGAGTTGG
HoxC11 ChIP F	CTCGCCTTCCCAAAATTTCC
HoxC11 ChIP R	AGCAGAAGTTGCCCAGGTTG
Hoxb13 ChIP F	TTCGAGCTGGGAGCGATTTA
Hoxb13 ChIP R	AGCCGAGGGTGAGGGTTCTA
Gata4 ChIP F	AAGAGCGCTTGCGTCTCTA
Gata4 ChIP R	TTGCTAGCCTCAGATCTACGG
Olig1 ChIP F	GGGTTACAGGCAGCCACCTA
Olig1 ChIP R	ATGCGGTGGAAGAGGATGAG
Bcor ChIP F	GTAAAACCGAAAGCGAGCAA
Bcor ChIP R	GAGGGTTTCTCCTCCGACTT
Fabp7 ChIP F	CTGGAGTCCCAGAGTTAGCC
Fabp7 ChIP R	TTTGCATTCCAAAGAAAGTCA
Polm ChIP F	TGACGGGCACAATTACACCA
Polm ChIP R	AAAGGCTTCCGCGTCCTAGA
Fzd1 ChIP F	ACATGAGCCCGTAAACCTTG
Fzd1 ChIP R	GGTGCCCTCCTACCTCAACT
Lhx5 prom ChIP F	CGGGAAGGCAAGCTATACTG
Lhx5 prom ChIP R	TCCCAAGGAGCCAGAGTAGA
Lhx5 PE ChIP F	ATGCCTCTCTGGCTTTTCAA
Lhx5 PE ChIP R	ATCAGCCCTAAGGTGGCTTT

Otx2 qPCR F	CATGATGTCTTATCTAAAGCAACCG
Otx2 qPCR R	GTCGAGCTGTGCCCTAGTA
Gapdh qPCR F	CATGGCCTTCCGTGTTCCT
Gapdh qPCR R	GCGGCACGTCAGATCCA
HoxB13 qPCR F	CATTCTGGAAAGCAGCGTTTG
HoxB13 qPCR R	TGTTGGCTGCATACTCCCG
Gata4 qPCR F	TTCCTCTCCCAGGAACATCAAA
Gata4 qPCR R	GCTGCACAACTGGGCTCTACTT
Gata6 qPCR F	TGCAAGATTGCATCATGACAGA
Gata6 qPCR R	TGACCTCAGATCAGCCACGTTA
FoxA2 qPCR F	CCATCAGCCCCACAAAATG
FoxA2 qPCR R	CCAAGCTGCCTGGCATG
Meis2 qPCR F	AAGTTGGGCAGCTTTCCTCA
Meis2 qPCR R	CCCCCAAGCTTTGAGTTCCT
Pax6 qPCR F	GATTGAGGCTCTGGAGAAAG
Pax6 qPCR R	CCATTTGGCCCTTCGATTAG
Fabp7 qPCR F	CGGACACAATGCACATTCAAG
Fabp7 qPCR R	CAACCGAACCACAGACTTACA
Ezh1 qPCR F	CGTCCATGTTTCCTGAGAATGG
Ezh1 qPCR R	GATGTTGGGTGTGCACTGAG
Ezh2 qPCR F	GGCACCGCAGAAGAACTGAA
Ezh2 qPCR R	GGCATTTGGTCCATCGATGTTTG
Eed qPCR F	CAGCCACCCTCTATTAGCAG
Eed qPCR R	GCATTTCCATGGCCAACATAG
Suz12 qPCR F	GGAAGTCCTGCTTGTGAAAG
Suz12 qPCR R	GATTGAGGTCAGGATTCAAAGG

Figure 9.3 Table of primers used in ChIP-qPCR and RT-qPCR experiments

Bibliography

Antonysamy, S., Condon, B., Druzina, Z., Bonanno, J. B., Gheyi, T., Zhang, F., MacEwan, I., Zhang, A., Ashok, S., Rodgers, L., Russell, M. and Gately Luz, J. (2013). Structural Context of Disease-Associated Mutations and Putative Mechanism of Autoinhibition Revealed by X-Ray Crystallographic Analysis of the EZH2-SET Domain. PLoS ONE *8*, e84147.

Arai, Y. (2005). Role of Fabp7, a Downstream Gene of Pax6, in the Maintenance of Neuroepithelial Cells during Early Embryonic Development of the Rat Cortex. Journal of Neuroscience *25*, 9752–9761.

Ardehali, M. B., Anselmo, A., Cochrane, J. C., Kundu, S., Sadreyev, R. I. and Kingston, R. E. (2017). Polycomb Repressive Complex 2 Methylates Elongin A to Regulate Transcription. Molecular Cell *68*, 872–884.e6.

Bach, C., Mueller, D., Buhl, S., Garcia-Cuellar, M. P. and Slany, R. K. (2009). Alterations of the CxxC domain preclude oncogenic activation of mixed-lineage leukemia 2. Oncogene *28*, 815–823.

Ballaré, C., Lange, M., Lapinaite, A., Martin, G. M., Morey, L., Pascual, G., Liefke, R., Simon, B., Shi, Y., Gozani, O., Carlomagno, T., Benitah, S. A. and Di Croce, L. (2012). Phf19 links methylated Lys36 of histone H3 to regulation of Polycomb activity. Nature Structural & Molecular Biology *19*, 1257–1265.

Bannister, A. J. and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. Cell Research *21*, 381–395.

Bantignies, F., Grimaud, C., Lavrov, S., Gabut, M. and Cavalli, G. (2003). Inheritance of polycomb-dependent chromosomal interactions in Drosophila. Genes and Development *17*, 2406–2420.

Bantignies, F., Roure, V., Comet, I., Leblanc, B., Schuettengruber, B., Bonnet, J., Tixier, V., Mas, A. and Cavalli, G. (2011). Polycomb-dependent regulatory contacts between distant hox loci in drosophila. Cell *144*, 214–226.

Barry, D. S., Pakan, J. M. and McDermott, K. W. (2014). Radial glial cells: Key organisers in CNS development. The International Journal of Biochemistry & Cell Biology *46*, 76–79.

Bartke, T., Vermeulen, M., Xhemalce, B., Robson, S. C., Mann, M. and Kouzarides, T. (2010). Nucleosome-Interacting Proteins Regulated by DNA and Histone Methylation. Cell *143*, 470–484.

Basnet, R., Gong, G. Q., Li, C. and Wang, M. W. (2018). Serum and glucocorticoid inducible protein kinases (SGKs): a potential target for cancer intervention. Acta Pharmaceutica Sinica B *8*, 767–771.

Beltran, M., Yates, C. M., Skalska, L., Dawson, M., Reis, F. P., Viiri, K., Fisher, C. L., Sibley, C. R., Foster, B. M., Bartke, T., Ule, J. and Jenner, R. G. (2016). The interaction of PRC2 with RNA or chromatin is mutually antagonistic. Genome Research *26*, 896–907.

Berger, S. L., Kouzarides, T., Shiekhattar, R. and Shilatifard, A. (2009). An operational definition of epigenetics. Genes & Development *23*, 781–783.

Beringer, M., Pisano, P., Di Carlo, V., Blanco, E., Chammas, P., Vizán, P., Gutiérrez, A., Aranda, S., Payer, B., Wierer, M. and Di Croce, L. (2016). EPOP Functionally Links Elongin and Polycomb in Pluripotent Stem Cells. Molecular Cell *64*, 645–658.

Bernstein, B. E., Mikkelsen, T. S., Xie, X., Kamal, M., Huebert, D. J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., Jaenisch, R., Wagschal, A., Feil, R., Schreiber, S. L. and Lander, E. S. (2006a). A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. Cell *125*, 315–326.

Bernstein, E., Duncan, E. M., Masui, O., Gil, J., Heard, E. and Allis, C. D. (2006b). Mouse Polycomb Proteins Bind Differentially to Methylated Histone H3 and RNA and Are Enriched in Facultative Heterochromatin. Molecular and Cellular Biology *26*, 2560–2569.

Bibel, M., Richter, J., Lacroix, E. and Barde, Y.-a. (2007). Generation of a defined and uniform population of CNS progenitors and neurons from mouse embryonic stem cells. Nature Protocols *2*, 1034–1043.

Bibel, M., Richter, J., Schrenk, K., Tucker, K. L., Staiger, V., Korte, M., Goetz, M. and Barde, Y.-a. (2004). Differentiation of mouse embryonic stem cells into a defined neuronal lineage. Nature Neuroscience *7*, 1003–1009.

Birke, M. (2002). The MT domain of the proto-oncoprotein MLL binds to CpG-containing DNA and discriminates against methylation. Nucleic Acids Research *30*, 958–965.

Blackledge, N. P., Farcas, A. M., Kondo, T., King, H. W., McGouran, J. F., Hanssen, L. L., Ito, S., Cooper, S., Kondo, K., Koseki, Y., Ishikura, T., Long, H. K., Sheahan, T. W., Brockdorff, N., Kessler, B. M., Koseki, H. and Klose, R. J. (2014). Variant PRC1 Complex-Dependent H2A Ubiquitylation Drives PRC2 Recruitment and Polycomb Domain Formation. Cell *157*, 1445–1459.

Blackledge, N. P., Fursova, N. A., Kelley, J. R., Huseyin, M. K., Feldmann, A. and Klose, R. J. (2019). PRC1 Catalytic Activity Is Central to Polycomb System Function. Molecular Cell , 1–18.

Bodor, C., Grossmann, V., Popov, N., Okosun, J., O'Riain, C., Tan, K., Marzec, J., Araf, S., Wang, J., Lee, A. M., Clear, A., Montoto, S., Matthews, J., Iqbal, S., Rajnai, H., Rosenwald, A., Ott, G., Campo, E., Rimsza, L. M., Smeland, E. B., Chan, W. C., Braziel, R. M., Staudt, L. M., Wright, G., Lister, T. A., Elemento, O., Hills, R., Gribben, J. G., Chelala, C., Matolcsy, A., Kohlmann, A., Haferlach, T., Gascoyne, R. D. and Fitzgibbon, J. (2013). EZH2 mutations are frequent and represent an early event in follicular lymphoma. Blood *122*, 3165–3168.

Boyer, L. A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L. A., Lee, T. I., Levine, S. S., Wernig, M., Tajonar, A., Ray, M. K., Bell, G. W., Otte, A. P., Vidal, M., Gifford, D. K., Young, R. A. and Jaenisch, R. (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature *441*, 349–353.

Bracken, A. P. (2003). EZH2 is downstream of the pRB-E2F pathway, essential for proliferation and amplified in cancer. The EMBO Journal *22*, 5323–5335.

Bracken, A. P. (2006). Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. Genes & Development *20*, 1123–1136.

Bracken, A. P., Kleine-Kohlbrecher, D., Dietrich, N., Pasini, D., Gargiulo, G., Beekman, C., Theilgaard-Monch, K., Minucci, S., Porse, B. T., Marine, J.-C., Hansen, K. H. and Helin, K. (2007). The Polycomb group proteins bind throughout the INK4A-ARF locus and are disassociated in senescent cells. Genes & Development *21*, 525–530.

Brien, G. L., Gambero, G., O'Connell, D. J., Jerman, E., Turner, S. A., Egan, C. M., Dunne, E. J., Jurgens, M. C., Wynne, K., Piao, L., Lohan, A. J., Ferguson, N., Shi, X., Sinha, K. M., Loftus, B. J., Cagney, G. and Bracken, A. P. (2012). Polycomb PHF19 binds H3K36me3 and recruits PRC2 and demethylase NO66 to embryonic stem cell genes during differentiation. Nature Structural & Molecular Biology *19*, 1273–1281.

Brinkman, A. B., Gu, H., Bartels, S. J., Zhang, Y., Matarese, F., Simmer, F., Marks, H., Bock, C., Gnirke, A., Meissner, A. and Stunnenberg, H. G. (2012). Sequential ChIP-bisulfite sequencing enables direct genome-scale investigation of chromatin and DNA methylation cross-talk. Genome Research *22*, 1128–1138.

Brookes, E., de Santiago, I., Hebenstreit, D., Morris, K. J., Carroll, T., Xie, S. Q., Stock, J. K., Heidemann, M., Eick, D., Nozaki, N., Kimura, H., Ragoussis, J., Teichmann, S. A. and Pombo, A. (2012). Polycomb Associates Genome-wide with a Specific RNA Polymerase II Variant, and Regulates Metabolic Genes in ESCs. Cell Stem Cell *10*, 157–170.

Burlingame, R., Love, W., Wang, B., Hamlin, R., Nguyen, H. and Moudrianakis, E. (1985). Crystallographic structure of the octameric histone core of the nucleosome at a resolution of 3.3 A. Science *228*, 546–553.

Cai, L., Rothbart, S. B., Lu, R., Xu, B., Chen, W.-Y., Tripathy, A., Rockowitz, S., Zheng, D., Patel, D. J., Allis, C. D., Strahl, B. D., Song, J. and Wang, G. G. (2013). An H3K36 Methylation-Engaging Tudor Motif of Polycomb-like Proteins Mediates PRC2 Complex Targeting. Molecular Cell *49*, 571–582.

Calebiro, D., Grassi, E. S., Eszlinger, M., Ronchi, C. L., Godbole, A., Bathon, K., Guizzardi, F., De Filippis, T., Krohn, K., Jaeschke, H., Schwarzmayr, T., Bircan, R., Gozu, H. I., Sancak, S., Niedziela, M., Strom, T. M., Fassnacht, M., Persani, L. and Paschke, R. (2016). Recurrent EZH1 mutations are a second hit in autonomous thyroid adenomas. Journal of Clinical Investigation *126*, 3383–3388.

Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R. S. and Zhang, Y. (2002). Role of Histone H3 Lysine 27 Methylation in Polycomb-Group Silencing. Science *298*, 1039–1043.

Cao, R. and Zhang, Y. (2004). SUZ12 is required for both the histone methyltransferase activity and the silencing function of the EED-EZH2 complex. Molecular Cell *15*, 57–67.

Casanova, M., Preissner, T., Cerase, A., Poot, R., Yamada, D., Li, X., Appanah, R., Bezstarosti, K., Demmers, J., Koseki, H. and Brockdorff, N. (2011). Polycomblike 2 facilitates the recruitment of PRC2 Polycomb group complexes to the inactive X chromosome and to target loci in embryonic stem cells. Development *138*, 1471–1482.

Cerase, A., Smeets, D., Tang, Y. A., Gdula, M., Kraus, F., Spivakov, M., Moindrot, B., Leleu, M., Tattermusch, A., Demmerle, J., Nesterova, T. B., Green, C., Otte, A. P., Schermelleh, L. and Brockdorff, N. (2014). Spatial separation of Xist RNA and polycomb proteins revealed by superresolution microscopy. Proceedings of the National Academy of Sciences *111*, 2235–2240.

Chamberlain, S. J., Yee, D. and Magnuson, T. (2008). Polycomb Repressive Complex 2 Is Dispensable for Maintenance of Embryonic Stem Cell Pluripotency. Stem Cells *26*, 1496–1505.

Champagne, K. and Kutateladze, T. (2009). Structural Insight Into Histone Recognition by the ING PHD Fingers. Current Drug Targets *10*, 432–441.

Chen, S., Jiao, L., Shubbar, M., Yang, X. and Liu, X. (2018). Unique Structural Platforms of Suz12 Dictate Distinct Classes of PRC2 for Chromatin Binding. Molecular Cell *69*, 840–852.e5.

Ciferri, C., Lander, G. C., Maiolica, A., Herzog, F., Aebersold, R. and Nogales, E. (2012). Molecular architecture of human polycomb repressive complex 2. eLife *1*, 1–22.

Cifuentes-Rojas, C., Hernandez, A. J., Sarma, K. and Lee, J. T. (2014). Regulatory Interactions between RNA and Polycomb Repressive Complex 2. Molecular Cell *55*, 171–185.

Cloos, P. A., Christensen, J., Agger, K. and Helin, K. (2008). Erasing the methyl mark: histone demethylases at the center of cellular differentiation and disease. Genes & Development *22*, 1115–1140.

Clouaire, T., Webb, S., Skene, P., Illingworth, R., Kerr, A., Andrews, R., Lee, J.-H., Skalnik, D. and Bird, A. (2012). Cfp1 integrates both CpG content and gene activity for accurate H3K4me3 deposition in embryonic stem cells. Genes & Development *26*, 1714–1728.

Collinson, A., Collier, A. J., Morgan, N. P., Sienerth, A. R., Chandra, T., Andrews, S. and Rugg-Gunn, P. J. (2016). Deletion of the Polycomb-Group Protein EZH2 Leads to Compromised Self-Renewal and Differentiation Defects in Human Embryonic Stem Cells. Cell Reports *17*, 2700–2714.

Conway, E., Jerman, E., Healy, E., Ito, S., Holoch, D., Oliviero, G., Deevy, O., Glancy, E., Fitzpatrick, D. J., Mucha, M., Watson, A., Rice, A. M., Chammas, P., Huang, C., Pratt-Kelly, I., Koseki, Y., Nakayama, M., Ishikura, T., Streubel, G., Wynne, K., Hokamp, K., McLysaght, A., Ciferri, C., Di Croce, L., Cagney, G., Margueron, R., Koseki, H. and Bracken, A. P. (2018). A Family of Vertebrate-Specific Polycombs Encoded by the LCOR/LCORL Genes Balance PRC2 Subtype Activities. Molecular Cell *70*, 408–421.e8.

Cooper, S., Dienstbier, M., Hassan, R., Schermelleh, L., Sharif, J., Blackledge, N. P., De Marco, V., Elderkin, S., Koseki, H., Klose, R., Heger, A. and Brockdorff, N. (2014). Targeting Polycomb to Pericentric Heterochromatin in Embryonic Stem Cells Reveals a Role for H2AK119u1 in PRC2 Recruitment. Cell Reports *7*, 1456–1470.

Cooper, S., Grijzenhout, A., Underwood, E., Ancelin, K., Zhang, T., Nesterova, T. B., Anil-Kirmizitas, B., Bassett, A., Kooistra, S. M., Agger, K., Helin, K., Heard, E. and Brockdorff, N. (2016). Jarid2 binds mono-ubiquitylated H2A lysine 119 to mediate crosstalk between Polycomb complexes PRC1 and PRC2. Nature Communications *7*, 13661.

Coré, N., Bel, S., Gaunt, S. J., Aurrand-Lions, M., Pearce, J., Fisher, A. and Djabali, M. (1997). Altered cellular proliferation and mesoderm patterning in Polycomb-M33-deficient mice. Development (Cambridge, England) *124*, 721–9.

Creyghton, M. P., Cheng, A. W., Welstead, G. G., Kooistra, T., Carey, B. W., Steine, E. J., Hanna, J., Lodato, M. A., Frampton, G. M., Sharp, P. A., Boyer, L. A., Young, R. A. and Jaenisch, R. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proceedings of the National Academy of Sciences *107*, 21931–21936.

Creyghton, M. P., Markoulaki, S., Levine, S. S., Hanna, J., Lodato, M. A., Sha, K., Young, R. A., Jaenisch, R. and Boyer, L. A. (2008). H2AZ Is Enriched at Polycomb Complex Target Genes in ES Cells and Is Necessary for Lineage Commitment. Cell *135*, 649–661.

Cruz-Molina, S., Respuela, P., Tebartz, C., Kolovos, P., Nikolic, M., Fueyo, R., van Ijcken, W. F., Grosveld, F., Frommolt, P., Bazzi, H. and Rada-Iglesias, A. (2017). PRC2 Facilitates the Regulatory Topology Required for Poised Enhancer Function during Pluripotent Stem Cell Differentiation. Cell Stem Cell *20*, 689–705.e9.

Czermin, B., Melfi, R., McCabe, D., Seitz, V., Imhof, A. and Pirrotta, V. (2002). Drosophila Enhancer of Zeste/ESC Complexes Have a Histone H3 Methyltransferase Activity that Marks Chromosomal Polycomb Sites. Cell *111*, 185–196.

da Rocha, S. T., Boeva, V., Escamilla-Del-Arenal, M., Ancelin, K., Granier, C., Matias, N. R., Sanulli, S., Chow, J., Schulz, E., Picard, C., Kaneko, S., Helin, K., Reinberg, D., Stewart, A. F., Wutz, A., Margueron, R. and Heard, E. (2014). Jarid2 Is Implicated in the Initial Xist-Induced Targeting of PRC2 to the Inactive X Chromosome. Molecular Cell *53*, 301–316.

Davidovich, C., Goodrich, K. J., Gooding, A. R. and Cech, T. R. (2014). A dimeric state for PRC2. Nucleic Acids Research *42*, 9236–9248.

Davidovich, C., Wang, X., Cifuentes-Rojas, C., Goodrich, K. J., Gooding, A. R., Lee, J. T. and Cech, T. R. (2015). Toward a consensus on the binding specificity and promiscuity of PRC2 for RNA. Molecular Cell *57*, 552–558.

Deckard, C. E. and Sczepanski, J. T. (2018). Polycomb repressive complex 2 binds RNA irrespective of stereochemistry. Chemical Communications *54*, 12061–12064.

del Mar Lorente, M., Marcos-Gutierrez, C., Perez, C., Schoorlemmer, J., Ramirez, A., Magin, T. and Vidal, M. (2000). Loss- and gain-of-function mutations show a Polycomb group function for Ring1A in mice. Development *127*, 5093–5100.

Denholtz, M., Bonora, G., Chronis, C., Splinter, E., de Laat, W., Ernst, J., Pellegrini, M. and Plath, K. (2013). Long-Range Chromatin Contacts in Embryonic Stem Cells Reveal a Role for Pluripotency Factors and Polycomb Proteins in Genome Organization. Cell Stem Cell *13*, 602–616.

Denissov, S., Hofemeister, H., Marks, H., Kranz, A., Ciotta, G., Singh, S., Anastassiadis, K., Stunnenberg, H. G. and Stewart, A. F. (2014). Mll2 is required for H3K4 trimethylation on bivalent promoters in embryonic stem cells, whereas Mll1 is redundant. Development *141*, 526–537.

Dietrich, N., Lerdrup, M., Landt, E., Agrawal-Singh, S., Bak, M., Tommerup, N., Rappsilber, J., Södersten, E. and Hansen, K. (2012). RESTMediated Recruitment

of Polycomb Repressor Complexes in Mammalian Cells. PLoS Genetics *8*, e1002494.

Dor, Y. and Cedar, H. (2018). Principles of DNA methylation and their implications for biology and medicine. The Lancet *392*, 777–786.

Dorighi, K. M., Swigut, T., Henriques, T., Bhanu, N. V., Scruggs, B. S., Nady, N., Still, C. D., Garcia, B. A., Adelman, K. and Wysocka, J. (2017). Mll3 and Mll4 Facilitate Enhancer RNA Synthesis and Transcription from Promoters Independently of H3K4 Monomethylation. Molecular Cell *66*, 568–576.e4.

Du, Q., Luu, P.-L., Stirzaker, C. and Clark, S. J. (2015). Methyl-CpG-binding domain proteins: readers of the epigenome. Epigenomics *7*, 1051–1073.

Duboule, D. and Dollé, P. (1989). The structural and functional organization of the murine HOX gene family resembles that of Drosophila homeotic genes. The EMBO Journal *8*, 1497–1505.

Dupont, C., Armant, D. and Brenner, C. (2009). Epigenetics: Definition, Mechanisms and Clinical Perspective. Seminars in Reproductive Medicine *27*, 351–357.

Eckert, D., Buhl, S., Weber, S., Jäger, R. and Schorle, H. (2005). The AP-2 family of transcription factors. Genome Biology *6*.

Edelstein, A. D., Tsuchida, M. A., Amodaj, N., Pinkard, H., Vale, R. D. and Stuurman, N. (2014). Advanced methods of microscope control using μManager software. Journal of Biological Methods *1*, 10.

Ernst, T., Chase, A. J., Score, J., Hidalgo-Curtis, C. E., Bryant, C., Jones, A. V., Waghorn, K., Zoi, K., Ross, F. M., Reiter, A., Hochhaus, A., Drexler, H. G., Duncombe, A., Cervantes, F., Oscier, D., Boultwood, J., Grand, F. H. and Cross, N. C. P. (2010). Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid disorders. Nature Genetics *42*, 722–726.

Eskeland, R., Leeb, M., Grimes, G. R., Kress, C., Boyle, S., Sproul, D., Gilbert, N., Fan, Y., Skoultchi, A. I., Wutz, A. and Bickmore, W. A. (2010). Ring1B Compacts Chromatin Structure and Represses Gene Expression Independent of Histone Ubiquitination. Molecular Cell *38*, 452–464.

Farcas, A. M., Blackledge, N. P., Sudbery, I., Long, H. K., McGouran, J. F., Rose, N. R., Lee, S., Sims, D., Cerase, A., Sheahan, T. W., Koseki, H., Brockdorff, N., Ponting, C. P., Kessler, B. M. and Klose, R. J. (2012). KDM2B links the Polycomb Repressive Complex 1 (PRC1) to recognition of CpG islands. eLife *1*, 1–26.

Faust, C., Lawson, K. A., Schork, N. J., Thiel, B. and Magnuson, T. (1998). The Polycomb-group gene eed is required for normal morphogenetic movements during gastrulation in the mouse embryo. Development (Cambridge, England) *125*, 4495–506.

Faust, C., Schumacher, A., Holdener, B. and Magnuson, T. (1995). The eed mutation disrupts anterior mesoderm production in mice. Development (Cambridge, England) *121*, 273–85.

Fazzio, T. G., Huff, J. T. and Panning, B. (2008). An RNAi Screen of Chromatin Proteins Identifies Tip60-p400 as a Regulator of Embryonic Stem Cell Identity. Cell *134*, 162–174.

Felsenfeld, G. (2014). A Brief History of Epigenetics. Cold Spring Harbor Perspectives in Biology *6*, a018200–a018200.

Ferrari, K. J., Scelfo, A., Jammula, S., Cuomo, A., Barozzi, I., Stützer, A., Fischle, W., Bonaldi, T. and Pasini, D. (2014). Polycomb-Dependent H3K27me1 and H3K27me2 Regulate Active Transcription and Enhancer Fidelity. Molecular Cell *53*, 49–62.

Fischle, W. (2003). Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. Genes & Development *17*, 1870–1881.

Francis, N. J., Kingston, R. E. and Woodcock, C. L. (2004). Chromatin Compaction by a Polycomb Group Protein Complex. Science *306*, 1574–1577.

Gammill, L. S. and Sive, H. (2001). otx2 Expression in the Ectoderm Activates Anterior Neural Determination and Is Required for Xenopus Cement Gland Formation. Developmental Biology *240*, 223–236.

Gao, Z., Zhang, J., Bonasio, R., Strino, F., Sawai, A., Parisi, F., Kluger, Y. and Reinberg, D. (2012). PCGF Homologs, CBX Proteins, and RYBP Define Functionally Distinct PRC1 Family Complexes. Molecular Cell *45*, 344–356.

Gentile, C., Berlivet, S., Mayran, A., Paquette, D., Guerard-Millet, F., Bajon, E., Dostie, J. and Kmita, M. (2019). PRC2-Associated Chromatin Contacts in the Developing Limb Reveal a Possible Mechanism for the Atypical Role of PRC2 in HoxA Gene Expression. Developmental Cell *50*, 1–13.

Glaser, S., Schaft, J., Lubitz, S., Vintersten, K., van der Hoeven, F., Tufftteland, K. R., Aasland, R., Anastassiadis, K., Ang, S. L. and Stewart, A. F. (2006). Multiple epigenetic maintenance factors implicated by the loss of MII2 in mouse development. Development *133*, 1423–1432.

Grijzenhout, A., Godwin, J., Koseki, H., Gdula, M. R., Szumska, D., McGouran, J. F., Bhattacharya, S., Kessler, B. M., Brockdorff, N. and Cooper, S. (2016). Functional analysis of AEBP2, a PRC2 Polycomb protein, reveals a Trithorax phenotype in embryonic development and in ESCs. Development *143*, 2716–2723.

Harms, P. W., Hristov, A. C., Kim, D. S., Anens, T., Quist, M. J., Siddiqui, J., Carskadon, S., Mehra, R., Fullen, D. R., Johnson, T. M., Chinnaiyan, A. M. and Palanisamy, N. (2014). Activating mutations of the oncogene EZH2 in cutaneous

melanoma revealed by next generation sequencing. Human Pathology: Case Reports *1*, 21–28.

Haub, O. and Goldfarb, M. (1991). Expression of the fibroblast growth factor-5 gene in the mouse embryo. Development *112*, 397–406.

Hauri, S., Comoglio, F., Seimiya, M., Gerstung, M., Glatter, T., Hansen, K., Aebersold, R., Paro, R., Gstaiger, M. and Beisel, C. (2016). A High-Density Map for Navigating the Human Polycomb Complexome. Cell reports *17*, 583–595.

He, A., Shen, X., Ma, Q., Cao, J., von Gise, A., Zhou, P., Wang, G., Marquez, V. E., Orkin, S. H. and Pu, W. T. (2012). PRC2 directly methylates GATA4 and represses its transcriptional activity. Genes and Development *26*, 37–42.

He, J., Shen, L., Wan, M., Taranova, O., Wu, H. and Zhang, Y. (2013). Kdm2b maintains murine embryonic stem cell status by recruiting PRC1 complex to CpG islands of developmental genes. Nature Cell Biology *15*, 373–384.

Herranz, N., Pasini, D., Diaz, V. M., Franci, C., Gutierrez, A., Dave, N., Escriva, M., Hernandez-Munoz, I., Di Croce, L., Helin, K., Garcia de Herreros, A. and Peiro, S. (2008). Polycomb Complex 2 Is Required for E-cadherin Repression by the Snail1 Transcription Factor. Molecular and Cellular Biology *28*, 4772–4781.

Herz, H.-M., Mohan, M., Garruss, A. S., Liang, K., Takahashi, Y.-h., Mickey, K., Voets, O., Verrijzer, C. P. and Shilatifard, A. (2012). Enhancer-associated H3K4 monomethylation by Trithorax-related, the Drosophila homolog of mammalian Mll3/Mll4. Genes & Development *26*, 2604–2620.

Heyting, C. (1996). Synaptonemal complexes: structure and function. Current Opinion in Cell Biology *8*, 389–396.

Hochedlinger, K., Yamada, Y., Beard, C. and Jaenisch, R. (2005). Ectopic Expression of Oct-4 Blocks Progenitor-Cell Differentiation and Causes Dysplasia in Epithelial Tissues. Cell *121*, 465–477.

Højfeldt, J. W., Laugesen, A., Willumsen, B. M., Damhofer, H., Hedehus, L., Tvardovskiy, A., Mohammad, F., Jensen, O. N. and Helin, K. (2018). Accurate H3K27 methylation can be established de novo by SUZ12-directed PRC2. Nature Structural & Molecular Biology *25*, 225–232.

Holoch, D. and Margueron, R. (2017). Polycomb Repressive Complex 2 Structure and Function. In Polycomb Group Proteins vol. 2, pp. 191–224. Elsevier.

Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A., Konermann, S., Agarwala, V., Li, Y., Fine, E. J., Wu, X., Shalem, O., Cradick, T. J., Marraffini, L. A., Bao, G. and Zhang, F. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. Nature Biotechnology *31*, 827–832.

Hu, D., Garruss, A. S., Gao, X., Morgan, M. A., Cook, M., Smith, E. R. and Shilatifard, A. (2013). The Mll2 branch of the COMPASS family regulates

bivalent promoters in mouse embryonic stem cells. Nature Structural & Molecular Biology *20*, 1093–1097.

Illingworth, R. S., Moffat, M., Mann, A. R., Read, D., Hunter, C. J., Pradeepa, M. M., Adams, I. R. and Bickmore, W. A. (2015). The E3 ubiquitin ligase activity of RING1B is not essential for early mouse development. Genes & Development *29*, 1897–1902.

Inoue, A., Jiang, L., Lu, F. and Zhang, Y. (2017). Genomic imprinting of Xist by maternal H3K27me3. Genes & Development *31*, 1927–1932.

Isono, K., Endo, T. A., Ku, M., Yamada, D., Suzuki, R., Sharif, J., Ishikura, T., Toyoda, T., Bernstein, B. E. and Koseki, H. (2013). SAM Domain Polymerization Links Subnuclear Clustering of PRC1 to Gene Silencing. Developmental Cell *26*, 565–577.

Jabbari, K. and Bernardi, G. (2004). Cytosine methylation and CpG, TpG (CpA) and TpA frequencies. Gene *333*, 143–149.

Jang, Y., Wang, C., Zhuang, L., Liu, C. and Ge, K. (2017). H3K4 Methyltransferase Activity Is Required for MLL4 Protein Stability. Journal of Molecular Biology *429*, 2046–2054.

Jankowska, A. M., Makishima, H., Tiu, R. V., Szpurka, H., Huang, Y., Traina, F., Visconte, V., Sugimoto, Y., Prince, C., O'Keefe, C., Hsi, E. D., List, A., Sekeres, M. A., Rao, A., McDevitt, M. A. and Maciejewski, J. P. (2011). Mutational spectrum analysis of chronic myelomonocytic leukemia includes genes associated with epigenetic regulation: UTX, EZH2, and DNMT3A. Blood *118*, 3932 LP – 3941.

Jenuwein, T. and Allis, C. D. (2001). Translating the histone code.

Jermann, P., Hoerner, L., Burger, L. and Schübeler, D. (2014). Short sequences can efficiently recruit histone H3 lysine 27 trimethylation in the absence of enhancer activity and DNA methylation. Proceedings of the National Academy of Sciences of the United States of America *111*, E3415–E3421.

Jiao, L. and Liu, X. (2015). Structural basis of histone H3K27 trimethylation by an active polycomb repressive complex 2. Science *350*, aac4383–aac4383.

Jones, R. S. and Gelbart, W. M. (1990). Genetic analysis of the enhancer of zeste locus and its role in gene regulation in Drosophila melanogaster. Genetics *126*, 185–99.

Joshi, O., Wang, S.-Y., Kuznetsova, T., Atlasi, Y., Peng, T., Fabre, P. J., Habibi, E., Shaik, J., Saeed, S., Handoko, L., Richmond, T., Spivakov, M., Burgess, D. and Stunnenberg, H. G. (2015). Dynamic Reorganization of Extremely Long-Range Promoter-Promoter Interactions between Two States of Pluripotency. Cell Stem Cell *17*, 748–757.

Justin, N., Zhang, Y., Tarricone, C., Martin, S. R., Chen, S., Underwood, E., De Marco, V., Haire, L. F., Walker, P. A., Reinberg, D., Wilson, J. R. and Gamblin, S. J. (2016). Structural basis of oncogenic histone H3K27M inhibition of human polycomb repressive complex 2. Nature Communications *7*, 11316.

Kalb, R., Latwiel, S., Baymaz, H. I., Jansen, P. W. T. C., Müller, C. W., Vermeulen, M. and Müller, J. (2014). Histone H2A monoubiquitination promotes histone H3 methylation in Polycomb repression. Nature Structural & Molecular Biology *21*, 569–571.

Kaneko, S., Son, J., Bonasio, R., Shen, S. S. and Reinberg, D. (2014). Nascent RNA interaction keeps PRC2 activity poised and in check. Genes & Development *28*, 1983–1988.

Kaneko, S., Son, J., Shen, S. S., Reinberg, D. and Bonasio, R. (2013). PRC2 binds active promoters and contacts nascent RNAs in embryonic stem cells. Nature structural & molecular biology *20*, 1258–64.

Kanhere, A., Viiri, K., Araújo, C. C., Rasaiyaah, J., Bouwman, R. D., Whyte, W. A., Pereira, C. F., Brookes, E., Walker, K., Bell, G. W., Pombo, A., Fisher, A. G., Young, R. A. and Jenner, R. G. (2010). Short RNAs Are Transcribed from Repressed Polycomb Target Genes and Interact with Polycomb Repressive Complex-2. Molecular Cell *38*, 675–688.

Kasinath, V., Faini, M., Poepsel, S., Reif, D., Feng, X. A., Stjepanovic, G., Aebersold, R. and Nogales, E. (2018). Structures of human PRC2 with its cofactors AEBP2 and JARID2. Science *359*, 940–944.

Ketel, C. S., Andersen, E. F., Vargas, M. L., Suh, J., Strome, S. and Simon, J. A. (2005). Subunit Contributions to Histone Methyltransferase Activities of Fly and Worm Polycomb Group Complexes. Molecular and Cellular Biology *25*, 6857–6868.

Kim, E., Kim, M., Woo, D. H., Shin, Y., Shin, J., Chang, N., Oh, Y. T., Kim, H., Rheey, J., Nakano, I., Lee, C., Joo, K. M., Rich, J. N., Nam, D. H. and Lee, J. (2013). Phosphorylation of EZH2 Activates STAT3 Signaling via STAT3 Methylation and Promotes Tumorigenicity of Glioblastoma Stem-like Cells. Cancer Cell *23*, 839–852.

Kim, H., Kang, K. and Kim, J. (2009). AEBP2 as a potential targeting protein for Polycomb Repression Complex PRC2. Nucleic Acids Research *37*, 2940–2950.

Kinkley, S., Helmuth, J., Polansky, J. K., Dunkel, I., Gasparoni, G., Fröhler, S., Chen, W., Walter, J., Hamann, A. and Chung, H. R. (2016). ReChIP-seq reveals widespread bivalency of H3K4me3 and H3K27me3 in CD4+ memory T cells. Nature Communications *7*, 12514.

Kipp, D. R., Quinn, C. M. and Fortin, P. D. (2013). Enzyme-Dependent Lysine Deprotonation in EZH2 Catalysis. Biochemistry *52*, 6866–6878.

Kobayashi, T., Deak, M., Morrice, N. and Cohen, P. (1999). Characterization of the structure and regulation of two novel isoforms of serum- and glucocorticoid-induced protein kinase. Biochemical Journal *344*, 189–197.

Kondo, T., Isono, K., Kondo, K., Endo, T. A., Itohara, S., Vidal, M. and Koseki, H. (2014). Polycomb Potentiates Meis2 Activation in Midbrain by Mediating Interaction of the Promoter with a Tissue-Specific Enhancer. Developmental Cell *28*, 94–101.

Kouzarides, T. (2007). Chromatin Modifications and Their Function. Cell *128*, 693–705.

Ku, M., Koche, R. P., Rheinbay, E., Mendenhall, E. M., Endoh, M., Mikkelsen, T. S., Presser, A., Nusbaum, C., Xie, X., Chi, A. S., Adli, M., Kasif, S., Ptaszek, L. M., Cowan, C. A., Lander, E. S., Koseki, H. and Bernstein, B. E. (2008). Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. PLoS Genetics *4*.

Kumar, S., Chinnusamy, V. and Mohapatra, T. (2018). Epigenetics of Modified DNA Bases: 5-Methylcytosine and Beyond. Frontiers in Genetics *9*, 1–14.

Kundu, S., Ji, F., Sunwoo, H., Jain, G., Lee, J. T., Sadreyev, R. I., Dekker, J. and Kingston, R. E. (2017). Polycomb Repressive Complex 1 Generates Discrete Compacted Domains that Change during Differentiation. Molecular Cell *65*, 432–446.e5.

Kuzmichev, A. (2002). Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. Genes & Development *16*, 2893–2905.

Kwon, T. (2003). Mechanism of histone lysine methyl transfer revealed by the structure of SET7/9-AdoMet. The EMBO Journal *22*, 292–303.

Lagarou, A., Mohd-Sarip, A., Moshkin, Y. M., Chalkley, G. E., Bezstarosti, K., Demmers, J. A. and Verrijzer, C. P. (2008). dKDM2 couples histone H2A ubiquitylation to histone H3 demethylation during Polycomb group silencing. Genes & Development *22*, 2799–2810.

Laible, G., Wolf, A., Dorn, R., Reuter, G., Nislow, C., Lebersorger, A., Popkin, D., Pillus, L. and Jenuwein, T. (1997). Mammalian homologues of the Polycomb-group gene Enhancer of zeste mediate gene silencing in Drosophila heterochromatin and at S. cerevisiae telomeres. The EMBO journal *16*, 3219–32.

Lau, M. S., Schwartz, M. G., Kundu, S., Savol, A. J., Wang, P. I., Marr, S. K., Grau, D. J., Schorderet, P., Sadreyev, R. I., Tabin, C. J. and Kingston, R. E. (2017). Mutation of a nucleosome compaction region disrupts Polycomb-mediated axial patterning. Science *355*, 1081–1084.

Lauberth, S. M., Nakayama, T., Wu, X., Ferris, A. L., Tang, Z., Hughes, S. H. and Roeder, R. G. (2013). H3K4me3 interactions with TAF3 regulate preinitiation complex assembly and selective gene activation. Cell *152*, 1021–1036.

Lavarone, E., Barbieri, C. M. and Pasini, D. (2019). Dissecting the role of H3K27 acetylation and methylation in PRC2 mediated control of cellular identity. Nature Communications *10*, 1679.

Lee, C.-h., Holder, M., Grau, D., Saldaña-Meyer, R., Yu, J.-R., Ganai, R. A., Zhang, J., Wang, M., LeRoy, G., Dobenecker, M.-W., Reinberg, D. and Armache, K.-J. (2018). Distinct Stimulatory Mechanisms Regulate the Catalytic Activity of Polycomb Repressive Complex 2. Molecular Cell *70*, 435–448.e5.

Lee, T. I., Jenner, R. G., Boyer, L. A., Guenther, M. G., Levine, S. S., Kumar, R. M., Chevalier, B., Johnstone, S. E., Cole, M. F., Isono, K.-i., Koseki, H., Fuchikami, T., Abe, K., Murray, H. L., Zucker, J. P., Yuan, B., Bell, G. W., Herbolsheimer, E., Hannett, N. M., Sun, K., Odom, D. T., Otte, A. P., Volkert, T. L., Bartel, D. P., Melton, D. A., Gifford, D. K., Jaenisch, R. and Young, R. A. (2006). Control of Developmental Regulators by Polycomb in Human Embryonic Stem Cells. Cell *125*, 301–313.

Lewis, E. B. (1978). A gene complex controlling segmetation in Drosophila. Nature *276*, 565–570.

Li, G., Margueron, R., Ku, M., Chambon, P., Bernstein, B. E. and Reinberg, D. (2010). Jarid2 and PRC2, partners in regulating gene expression. Genes & Development *24*, 368–380.

Li, H., Liefke, R., Jiang, J., Kurland, J. V., Tian, W., Deng, P., Zhang, W., He, Q., Patel, D. J., Bulyk, M. L., Shi, Y. and Wang, Z. (2017). Polycomb-like proteins link the PRC2 complex to CpG islands. Nature *549*, 287–291.

Li, Q., Wang, X., Lu, Z., Zhang, B., Guan, Z., Liu, Z., Zhong, Q., Gu, L., Zhou, J., Zhu, B., Ji, J. and Deng, D. (2010). Polycomb CBX7 Directly Controls Trimethylation of Histone H3 at Lysine 9 at the p16 Locus. PLoS ONE *5*, e13732.

Liefke, R., Karwacki-Neisius, V. and Shi, Y. (2016). EPOP Interacts with Elongin BC and USP7 to Modulate the Chromatin Landscape. Molecular Cell *64*, 659–672.

Liefke, R. and Shi, Y. (2015). The PRC2-associated factor C17orf96 is a novel CpG island regulator in mouse ES cells. Cell Discovery *1*, 1–11.

Long, H. K., Blackledge, N. P. and Klose, R. J. (2013). ZF-CxxC domain-containing proteins, CpG islands and the chromatin connection. Biochemical Society Transactions *41*, 727–740.

Lowe, S. W. and Sherr, C. J. (2003). Tumor suppression by Ink4aArf: progress and puzzles. Current Opinion in Genetics & Development *13*, 77–83.

Lu, R., Yang, A. and Jin, Y. (2011). Dual Functions of T-Box 3 (Tbx3) in the Control of Self-renewal and Extraembryonic Endoderm Differentiation in Mouse Embryonic Stem Cells. Journal of Biological Chemistry *286*, 8425–8436.

Lubitz, S., Glaser, S., Schaft, J., Stewart, A. F. and Anastassiadis, K. (2007). Increased Apoptosis and Skewed Differentiation in Mouse Embryonic Stem Cells Lacking the Histone Methyltransferase Mll2. Molecular Biology of the Cell *18*, 2356–2366.

Luger, K., Rechsteiner, T. J., Flaus, A. J., Waye, M. M. and Richmond, T. J. (1997). Characterization of nucleosome core particles containing histone proteins made in bacteria 1 1Edited by A. Klug. Journal of Molecular Biology *272*, 301–311.

Margueron, R., Justin, N., Ohno, K., Sharpe, M. L., Son, J., Drury III, W. J., Voigt, P., Martin, S. R., Taylor, W. R., De Marco, V., Pirrotta, V., Reinberg, D. and Gamblin, S. J. (2009). Role of the polycomb protein EED in the propagation of repressive histone marks. Nature *461*, 762–767.

Margueron, R., Li, G., Sarma, K., Blais, A., Zavadil, J., Woodcock, C. L., Dynlacht, B. D. and Reinberg, D. (2008). Ezh1 and Ezh2 Maintain Repressive Chromatin through Different Mechanisms. Molecular Cell *32*, 503–518.

Martín Caballero, I., Hansen, J., Leaford, D., Pollard, S. and Hendrich, B. D. (2009). The Methyl-CpG Binding Proteins Mecp2, Mbd2 and Kaiso Are Dispensable for Mouse Embryogenesis, but Play a Redundant Function in Neural Differentiation. PLoS ONE *4*, e4315.

Mas, G., Blanco, E., Ballaré, C., Sansó, M., Spill, Y. G., Hu, D., Aoi, Y., Le Dily, F., Shilatifard, A., Marti-Renom, M. A. and Di Croce, L. (2018). Promoter bivalency favors an open chromatin architecture in embryonic stem cells. Nature Genetics *50*, 1452–1462.

McGinty, R. K., Henrici, R. C. and Tan, S. (2014). Crystal structure of the PRC1 ubiquitylation module bound to the nucleosome. Nature *514*, 591–596.

McHugh, C. A., Chen, C.-K., Chow, A., Surka, C. F., Tran, C., McDonel, P., Pandya-Jones, A., Blanco, M., Burghard, C., Moradian, A., Sweredoski, M. J., Shishkin, A. A., Su, J., Lander, E. S., Hess, S., Plath, K. and Guttman, M. (2015). The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. Nature *521*, 232–236.

Mendenhall, E. M., Koche, R. P., Truong, T., Zhou, V. W., Issac, B., Chi, A. S., Ku, M. and Bernstein, B. E. (2010). GC-rich sequence elements recruit PRC2 in mammalian ES cells. PLoS Genetics *6*, 1–10.

Mikkelsen, T. S., Ku, M., Jaffe, D. B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.-K., Koche, R. P., Lee, W., Mendenhall, E., O'Donovan, A., Presser, A., Russ, C., Xie, X., Meissner, A., Wernig, M.,

Jaenisch, R., Nusbaum, C., Lander, E. S. and Bernstein, B. E. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553–560.

Mohn, F., Weber, M., Rebhan, M., Roloff, T. C., Richter, J., Stadler, M. B., Bibel, M. and Schübeler, D. (2008). Lineage-Specific Polycomb Targets and De Novo DNA Methylation Define Restriction and Potential of Neuronal Progenitors. Molecular Cell *30*, 755–766.

Morey, L., Pascual, G., Cozzuto, L., Roma, G., Wutz, A., Benitah, S. A. and Di Croce, L. (2012). Nonoverlapping Functions of the Polycomb Group Cbx Family of Proteins in Embryonic Stem Cells. Cell Stem Cell *10*, 47–62.

Morin, R. D., Johnson, N. A., Severson, T. M., Mungall, A. J., An, J., Goya, R., Paul, J. E., Boyle, M., Woolcock, B. W., Kuchenbauer, F., Yap, D., Humphries, R. K., Griffith, O. L., Shah, S., Zhu, H., Kimbara, M., Shashkin, P., Charlot, J. F., Tcherpakov, M., Corbett, R., Tam, A., Varhol, R., Smailus, D., Moksa, M., Zhao, Y., Delaney, A., Qian, H., Birol, I., Schein, J., Moore, R., Holt, R., Horsman, D. E., Connors, J. M., Jones, S., Aparicio, S., Hirst, M., Gascoyne, R. D. and Marra, M. A. (2010). Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. Nature Genetics *42*, 181–185.

Motoyama, J., Kitajima, K., Kojima, M., Kondo, S. and Takeuchi, T. (1997). Organogenesis of the liver, thymus and spleen is affected in jumonji mutant mice. Mechanisms of Development *66*, 27–37.

Mozzetta, C., Pontis, J., Fritsch, L., Robin, P., Portoso, M., Proux, C., Margueron, R. and Ait-Si-Ali, S. (2014). The Histone H3 Lysine 9 Methyltransferases G9a and GLP Regulate Polycomb Repressive Complex 2-Mediated Gene Silencing. Molecular Cell *53*, 277–289.

Müller, H. and Helin, K. (2000). The E2F transcription factors: key regulators of cell proliferation. Biochimica et Biophysica Acta (BBA) - Reviews on Cancer *1470*, M1–M12.

Müller, J., Hart, C. M., Francis, N. J., Vargas, M. L., Sengupta, A., Wild, B., Miller, E. L., O'Connor, M. B., Kingston, R. E. and Simon, J. a. (2002). Histone methyltransferase activity of a Drosophila Polycomb group repressor complex. Cell *111*, 197–208.

Musselman, C. A., Avvakumov, N., Watanabe, R., Abraham, C. G., Lalonde, M.-E., Hong, Z., Allen, C., Roy, S., Nuñez, J. K., Nickoloff, J., Kulesza, C. A., Yasui, A., Côté, J. and Kutateladze, T. G. (2012). Molecular basis for H3K36me3 recognition by the Tudor domain of PHF1. Nature Structural & Molecular Biology *19*, 1266–1272.

Nekrasov, M., Klymenko, T., Fraterman, S., Papp, B., Oktaba, K., Köcher, T., Cohen, A., Stunnenberg, H. G., Wilm, M. and Müller, J. (2007). Pcl-PRC2

is needed to generate high levels of H3-K27 trimethylation at Polycomb target genes. The EMBO Journal *26*, 4078–4088.

Németh, A. and Längst, G. (2004). Chromatin higher order structure: Opening up chromatin for transcription. Briefings in Functional Genomics and Proteomics *2*, 334–343.

Nishimura, K., Fukagawa, T., Takisawa, H., Kakimoto, T. and Kanemaki, M. (2009). An auxin-based degron system for the rapid depletion of proteins in nonplant cells. Nature Methods *6*, 917–922.

Nora, E. P., Lajoie, B. R., Schulz, E. G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N. L., Meisig, J., Sedat, J., Gribnau, J., Barillot, E., Blüthgen, N., Dekker, J. and Heard, E. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature *485*, 381–385.

Nowialis, P., Lopusna, K., Opavska, J., Haney, S. L., Abraham, A., Sheng, P., Riva, A., Natarajan, A., Guryanova, O., Simpson, M., Hlady, R., Xie, M. and Opavsky, R. (2019). Catalytically inactive Dnmt3b rescues mouse embryonic development by accessory and repressive functions. Nature Communications *10*, 4374.

Obier, N., Lin, Q., Cauchy, P., Hornich, V., Zenke, M., Becker, M. and Müller, A. M. (2015). Polycomb Protein EED is Required for Silencing of Pluripotency Genes upon ESC Differentiation. Stem Cell Reviews and Reports *11*, 50–61.

O'Carroll, D., Erhardt, S., Pagani, M., Barton, S. C., Surani, M. A. and Jenuwein, T. (2001). The Polycomb-Group Gene Ezh2 Is Required for Early Mouse Development. Molecular and Cellular Biology *21*, 4330–4336.

Oksuz, O., Narendra, V., Lee, C.-h., Descostes, N., LeRoy, G., Raviram, R., Blumenberg, L., Karch, K., Rocha, P. P., Garcia, B. A., Skok, J. A. and Reinberg, D. (2018). Capturing the Onset of PRC2-Mediated Repressive Domain Formation. Molecular Cell *70*, 1149–1162.e5.

O'Loghlen, A., Muñoz-Cabello, A. M., Gaspar-Maia, A., Wu, H.-A., Banito, A., Kunowska, N., Racek, T., Pemberton, H. N., Beolchi, P., Lavial, F., Masui, O., Vermeulen, M., Carroll, T., Graumann, J., Heard, E., Dillon, N., Azuara, V., Snijders, A. P., Peters, G., Bernstein, E. and Gil, J. (2012). MicroRNA Regulation of Cbx7 Mediates a Switch of Polycomb Orthologs during ESC Differentiation. Cell Stem Cell *10*, 33–46.

Ott, H. M., Graves, A. P., Pappalardi, M. B., Huddleston, M., Halsey, W. S., Hughes, A. M., Groy, A., Dul, E., Jiang, Y., Bai, Y., Annan, R., Verma, S. K., Knight, S. D., Kruger, R. G., Dhanak, D., Schwartz, B., Tummino, P. J., Creasy, C. L. and McCabe, M. T. (2014). A687V EZH2 Is a Driver of Histone H3 Lysine 27 (H3K27) Hypertrimethylation. Molecular Cancer Therapeutics *13*, 3062–3073.

Pan, G., Tian, S., Nie, J., Yang, C., Ruotti, V., Wei, H., Jonsdottir, G. A., Stewart, R. and Thomson, J. A. (2007). Whole-Genome Analysis of Histone H3 Lysine 4 and Lysine 27 Methylation in Human Embryonic Stem Cells. Cell Stem Cell *1*, 299–312.

Pasini, D., Bracken, A. P., Hansen, J. B., Capillo, M. and Helin, K. (2007). The Polycomb Group Protein Suz12 Is Required for Embryonic Stem Cell Differentiation. Molecular and Cellular Biology *27*, 3769–3779.

Pasini, D., Bracken, A. P., Jensen, M. R., Denchi, E. L. and Helin, K. (2004). Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity. The EMBO Journal *23*, 4061–4071.

Pasini, D., Cloos, P. A. C., Walfridsson, J., Olsson, L., Bukowski, J.-P., Johansen, J. V., Bak, M., Tommerup, N., Rappsilber, J. and Helin, K. (2010). JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. Nature *464*, 306–310.

Peng, J. C., Valouev, A., Swigut, T., Zhang, J., Zhao, Y., Sidow, A. and Wysocka, J. (2009). Jarid2/Jumonji Coordinates Control of PRC2 Enzymatic Activity and Target Gene Occupancy in Pluripotent Cells. Cell *139*, 1290–1302.

Pengelly, A. R., Kalb, R., Finkl, K. and Müller, J. (2015). Transcriptional repression by PRC1 in the absence of H2A monoubiquitylation. Genes and Development *29*, 1487–1492.

Perino, M., van Mierlo, G., Karemaker, I. D., van Genesen, S., Vermeulen, M., Marks, H., van Heeringen, S. J. and Veenstra, G. J. C. (2018). MTF2 recruits Polycomb Repressive Complex 2 by helical-shape-selective DNA binding. Nature Genetics *50*, 1002–1010.

Peters, A. H. F. M., Kubicek, S., Mechtler, K., O'Sullivan, R. J., Derijck, A. A. H. A., Perez-Burgos, L., Kohlmaier, A., Opravil, S., Tachibana, M., Shinkai, Y., Martens, J. H. A. and Jenuwein, T. (2003). Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. Molecular cell *12*, 1577–89.

Poepsel, S., Kasinath, V. and Nogales, E. (2018). Cryo-EM structures of PRC2 simultaneously engaged with two functionally distinct nucleosomes. Nature Structural & Molecular Biology *25*, 154–162.

Qin, S. and Zhang, C.-L. (2012). Role of Kruppel-Like Factor 4 in Neurogenesis and Radial Neuronal Migration in the Developing Cerebral Cortex. Molecular and Cellular Biology *32*, 4297–4305.

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S. a., Flynn, R. a. and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. Nature *470*, 279–283.

Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A. and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. Nature Protocols *8*, 2281–2308.

Rando, O. J. (2012). Combinatorial complexity in chromatin structure and function: Revisiting the histone code.

Riising, E. M., Comet, I., Leblanc, B., Wu, X., Johansen, J. V. and Helin, K. (2014). Gene Silencing Triggers Polycomb Repressive Complex 2 Recruitment to CpG Islands Genome Wide. Molecular Cell *55*, 347–360.

Ringel, A. E., Cieniewicz, A. M., Taverna, S. D. and Wolberger, C. (2015). Nucleosome competition reveals processive acetylation by the SAGA HAT module. Proceedings of the National Academy of Sciences *112*, E5461–E5470.

Rinn, J. L., Kertesz, M., Wang, J. K., Squazzo, S. L., Xu, X., Brugmann, S. A., Goodnough, L. H., Helms, J. A., Farnham, P. J., Segal, E. and Chang, H. Y. (2007). Functional Demarcation of Active and Silent Chromatin Domains in Human HOX Loci by Noncoding RNAs. Cell *129*, 1311–1323.

Robinson, P. J. and Rhodes, D. (2006). Structure of the 30nm' chromatin fibre: A key role for the linker histone. Current Opinion in Structural Biology *16*, 336–343.

Ruthenburg, A. J., Li, H., Patel, D. J. and David Allis, C. (2007). Multivalent engagement of chromatin modifications by linked binding modules.

Sanulli, S., Justin, N., Teissandier, A., Ancelin, K., Portoso, M., Caron, M., Michaud, A., Lombard, B., da Rocha, S. T., Offer, J., Loew, D., Servant, N., Wassef, M., Burlina, F., Gamblin, S. J., Heard, E. and Margueron, R. (2015). Jarid2 Methylation via the PRC2 Complex Regulates H3K27me3 Deposition during Cell Differentiation. Molecular Cell *57*, 769–783.

Sarma, K., Margueron, R., Ivanov, A., Pirrotta, V. and Reinberg, D. (2008). Ezh2 Requires PHF1 To Efficiently Catalyze H3 Lysine 27 Trimethylation In Vivo. Molecular and Cellular Biology *28*, 2718–2731.

Schmitges, F. W., Prusty, A. B., Faty, M., Stützer, A., Lingaraju, G. M., Aiwazian, J., Sack, R., Hess, D., Li, L., Zhou, S., Bunker, R. D., Wirth, U., Bouwmeester, T., Bauer, A., Ly-Hartig, N., Zhao, K., Chan, H., Gu, J., Gut, H., Fischle, W., Müller, J. and Thomä, N. H. (2011). Histone Methylation by PRC2 Is Inhibited by Active Chromatin Marks. Molecular Cell *42*, 330–341.

Schmitz, S. U., Albert, M., Malatesta, M., Morey, L., Johansen, J. V., Bak, M., Tommerup, N., Abarrategui, I. and Helin, K. (2011). Jarid1b targets genes regulating development and is involved in neural differentiation. The EMBO Journal *30*, 4586–4600.

Schoeftner, S., Sengupta, A. K., Kubicek, S., Mechtler, K., Spahn, L., Koseki, H., Jenuwein, T. and Wutz, A. (2006). Recruitment of PRC1 function at the

initiation of X inactivation independent of PRC2 and silencing. The EMBO Journal *25*, 3110–3122.

Schoenfelder, S., Sugar, R., Dimond, A., Javierre, B.-m., Armstrong, H., Mifsud, B., Dimitrova, E., Matheson, L., Tavares-Cadete, F., Furlan-Magaril, M., Segonds-Pichon, A., Jurkowski, W., Wingett, S. W., Tabbada, K., Andrews, S., Herman, B., LeProust, E., Osborne, C. S., Koseki, H., Fraser, P., Luscombe, N. M. and Elderkin, S. (2015). Polycomb repressive complex PRC1 spatially constrains the mouse embryonic stem cell genome. Nature Genetics *47*, 1179–1186.

Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B. and Cavalli, G. (2007). Genome Regulation by Polycomb and Trithorax Proteins. Cell *128*, 735–745.

Sen, S., Block, K. F., Pasini, A., Baylin, S. B. and Easwaran, H. (2016). Genome-wide positioning of bivalent mononucleosomes. BMC Medical Genomics *9*, 1–14.

Shan, Y., Liang, Z., Xing, Q., Zhang, T., Wang, B., Tian, S., Huang, W., Zhang, Y., Yao, J., Zhu, Y., Huang, K., Liu, Y., Wang, X., Chen, Q., Zhang, J., Shang, B., Li, S., Shi, X., Liao, B., Zhang, C., Lai, K., Zhong, X., Shu, X., Wang, J., Yao, H., Chen, J., Pei, D. and Pan, G. (2017). PRC2 specifies ectoderm lineages and maintains pluripotency in primed but not naïve ESCs. Nature Communications *8*, 672.

Shao, Z., Raible, F., Mollaaghababa, R., Guyon, J. R., Wu, C.-t., Bender, W. and Kingston, R. E. (1999). Stabilization of Chromatin Structure by PRC1, a Polycomb Complex. Cell *98*, 37–46.

Shema, E., Jones, D., Shoresh, N., Donohue, L., Ram, O. and Bernstein, B. E. (2016). Single-molecule decoding of combinatorially modified nucleosomes. Science *352*, 717–721.

Shen, X., Kim, W., Fujiwara, Y., Simon, M. D., Liu, Y., Mysliwiec, M. R., Yuan, G.-C., Lee, Y. and Orkin, S. H. (2009). Jumonji Modulates Polycomb Activity and Self-Renewal versus Differentiation of Stem Cells. Cell *139*, 1303–1314.

Shen, X., Liu, Y., Hsu, Y.-J., Fujiwara, Y., Kim, J., Mao, X., Yuan, G.-C. and Orkin, S. H. (2008). EZH1 Mediates Methylation on Histone H3 Lysine 27 and Complements EZH2 in Maintaining Stem Cell Identity and Executing Pluripotency. Molecular Cell *32*, 491–502.

Shilatifard, A. (2012). The COMPASS Family of Histone H3K4 Methylases: Mechanisms of Regulation in Development and Disease Pathogenesis. Annual Review of Biochemistry *81*, 65–95.

Shogren-Knaak, M. (2006). Histone H4-K16 Acetylation Controls Chromatin Structure and Protein Interactions. Science *311*, 844–847.

Silva, J., Mak, W., Zvetkova, I., Appanah, R., Nesterova, T. B., Webster, Z., Peters, A. H., Jenuwein, T., Otte, A. P. and Brockdorff, N. (2003). Establishment of histone H3 methylation on the inactive X chromosome requires transient recruitment of Eed-Enx1 polycomb group complexes. Developmental Cell *4*, 481–495.

Simon, J., Chiang, A. and Bender, W. (1992). Ten different Polycomb group genes are required for spatial control of the abdA and AbdB homeotic products. Development *114*, 493–505.

Skourti-Stathaki, K., Proudfoot, N. J. and Gromak, N. (2011). Human Senataxin Resolves RNA/DNA Hybrids Formed at Transcriptional Pause Sites to Promote Xrn2-Dependent Termination. Molecular Cell *42*, 794–805.

Skourti-Stathaki, K., Torlai Triglia, E., Warburton, M., Voigt, P., Bird, A. and Pombo, A. (2019). R-Loops Enhance Polycomb Repression at a Subset of Developmental Regulator Genes. Molecular Cell *73*, 930–945.e4.

Sneeringer, C. J., Scott, M. P., Kuntz, K. W., Knutson, S. K., Pollock, R. M., Richon, V. M. and Copeland, R. A. (2010). Coordinated activities of wild-type plus mutant EZH2 drive tumor-associated hypertrimethylation of lysine 27 on histone H3 (H3K27) in human B-cell lymphomas. Proceedings of the National Academy of Sciences of the United States of America *107*, 20980–20985.

Son, J., Shen, S. S., Margueron, R. and Reinberg, D. (2013). Nucleosome-binding activities within JARID2 and EZH1 regulate the function of PRC2 on chromatin. Genes & Development *27*, 2663–2677.

Stock, J. K., Giadrossi, S., Casanova, M., Brookes, E., Vidal, M., Koseki, H., Brockdorff, N., Fisher, A. G. and Pombo, A. (2007). Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. Nature Cell Biology *9*, 1428–1435.

Struhl, G. and Akam, M. (1985). Altered distributions of Ultrabithorax transcripts in extra sex combs mutant embryos of Drosophila. The EMBO Journal *4*, 3259–3264.

Takahashi, K. and Yamanaka, S. (2006). Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. Cell *126*, 663–676.

Takebayashi-Suzuki, K., Konishi, H., Miyamoto, T., Nagata, T., Uchida, M. and Suzuki, A. (2018). Coordinated regulation of the dorsal-ventral and anterior-posterior patterning of Xenopus embryos by the BTB/POZ zinc finger protein Zbtb14. Development, Growth & Differentiation *60*, 158–173.

Takihara, Y., Tomotsune, D., Shirai, M., Katoh-Fukui, Y., Nishii, K., Motaleb, M. A., Nomura, M., Tsuchiya, R., Fujita, Y., Shibata, Y., Higashinakagawa, T. and Shimada, K. (1997). Targeted disruption of the mouse homologue of the

Drosophila polyhomeotic gene leads to altered anteroposterior patterning and neural crest defects. Development (Cambridge, England) *124*, 3673–82.

Tanay, A., O'Donnell, A. H., Damelin, M. and Bestor, T. H. (2007). Hyperconserved CpG domains underlie Polycomb-binding sites. Proceedings of the National Academy of Sciences *104*, 5521–5526.

Tavares, L., Dimitrova, E., Oxley, D., Webster, J., Poot, R., Demmers, J., Bezstarosti, K., Taylor, S., Ura, H., Koide, H., Wutz, A., Vidal, M., Elderkin, S. and Brockdorff, N. (2012). RYBP-PRC1 Complexes Mediate H2A Ubiquitylation at Polycomb Target Sites Independently of PRC2 and H3K27me3. Cell *148*, 664–678.

Taverna, S. D., Li, H., Ruthenburg, A. J., Allis, C. D. and Patel, D. J. (2007). How chromatin-binding modules interpret histone modifications: Lessons from professional pocket pickers.

Tie, F., Banerjee, R., Saiakhova, A. R., Howard, B., Monteith, K. E., Scacheri, P. C., Cosgrove, M. S. and Harte, P. J. (2014). Trithorax monomethylates histone H3K4 and interacts directly with CBP to promote H3K27 acetylation and antagonize Polycomb silencing. Development *141*, 1129–1139.

Tie, F., Banerjee, R., Stratton, C. A., Prasad-Sinha, J., Stepanik, V., Zlobin, A., Diaz, M. O., Scacheri, P. C. and Harte, P. J. (2009). CBP-mediated acetylation of histone H3 lysine 27 antagonizes Drosophila Polycomb silencing. Development *136*, 3131–3141.

Tiwari, V. K., Cope, L., McGarvey, K. M., Ohm, J. E. and Baylin, S. B. (2008a). A novel 6C assay uncovers Polycomb-mediated higher order chromatin conformations. Genome Research *18*, 1171–1179.

Tiwari, V. K., McGarvey, K. M., Licchesi, J. D., Ohm, J. E., Herman, J. G., Schübeler, D. and Baylin, S. B. (2008b). PcG Proteins, DNA Methylation, and Gene Repression by Chromatin Looping. PLoS Biology *6*, e306.

Tolhuis, B., Blom, M., Kerkhoven, R. M., Pagie, L., Teunissen, H., Nieuwland, M., Simonis, M., de Laat, W., van Lohuizen, M. and van Steensel, B. (2011). Interactions among Polycomb Domains Are Guided by Chromosome Architecture. PLoS Genetics *7*, e1001343.

Trojer, P. and Reinberg, D. (2007). Facultative Heterochromatin: Is There a Distinctive Molecular Signature? Molecular Cell *28*, 1–13.

Trouillas, M., Saucourt, C., Guillotin, B., Gauthereau, X., Ding, L., Buchholz, F., Doss, M., Sachinidis, A., Hescheler, J., Hummel, O., Huebner, N., Kolde, R., Vilo, J., Schulz, H. and Bœuf, H. (2009). Three LIF-dependent signatures and gene clusters with atypical expression profiles, identified by transcriptome studies in mouse ES cells and early derivatives. BMC Genomics *10*, 73.

Vermeulen, M., Mulder, K. W., Denissov, S., Pijnappel, W. W., van Schaik, F. M., Varier, R. A., Baltissen, M. P., Stunnenberg, H. G., Mann, M. and Timmers, H. T. M. (2007). Selective Anchoring of TFIID to Nucleosomes by Trimethylation of Histone H3 Lysine 4. Cell *131*, 58–69.

Vieux-Rochas, M., Fabre, P. J., Leleu, M., Duboule, D. and Noordermeer, D. (2015). Clustering of mammalian Hox genes with other H3K27me3 targets within an active nuclear domain. Proceedings of the National Academy of Sciences *112*, 4672–4677.

Voigt, P., LeRoy, G., Drury, W. J., Zee, B. M., Son, J., Beck, D. B., Young, N. L., Garcia, B. A. and Reinberg, D. (2012). Asymmetrically Modified Nucleosomes. Cell *151*, 181–193.

Voncken, J. W., Roelen, B. A. J., Roefs, M., de Vries, S., Verhoeven, E., Marino, S., Deschamps, J. and van Lohuizen, M. (2003). Rnf2 (Ring1b) deficiency causes gastrulation arrest and cell cycle inhibition. Proceedings of the National Academy of Sciences *100*, 2468–2473.

Wachter, E., Quante, T., Merusi, C., Arczewska, A., Stewart, F., Webb, S. and Bird, A. (2014). Synthetic CpG islands reveal DNA sequence determinants of chromatin structure. eLife *3*, e03397.

Waddington, C. H. (1968). Towards A Theoretical Biology. Nature *218*, 525–527.

Waddington, C. H. (2012). The Epigenotype. International Journal of Epidemiology *41*, 10–13.

Walker, E., Chang, W. Y., Hunkapiller, J., Cagney, G., Garcha, K., Torchia, J., Krogan, N. J., Reiter, J. F. and Stanford, W. L. (2010). Polycomb-like 2 Associates with PRC2 and Regulates Transcriptional Networks during Mouse Embryonic Stem Cell Self-Renewal and Differentiation. Cell Stem Cell *6*, 153–166.

Wang, D., Sang, H., Zhang, K., Nie, Y., Zhao, S., Zhang, Y., He, N., Wang, Y., Xu, Y., Xie, X., Li, Z. and Liu, N. (2017). Stat3 phosphorylation is required for embryonic stem cells ground state maintenance in 2i culture media. Oncotarget *8*, 31227–31237.

Wang, H., Wang, L., Erdjument-Bromage, H., Vidal, M., Tempst, P., Jones, R. S. and Zhang, Y. (2004). Role of histone H2A ubiquitination in Polycomb silencing. Nature *431*, 873–878.

Wang, X., Long, Y., Paucek, R. D., Gooding, A. R., Lee, T., Burdorf, R. M. and Cech, T. R. (2019). Regulation of histone methylation by automethylation of PRC2. Genes & Development *33*, 1416–1427.

Wei, Y., Mizzen, C. A., Cook, R. G., Gorovsky, M. A. and Allis, C. D. (1998). Phosphorylation of histone H3 at serine 10 is correlated with chromosome condensation during mitosis and meiosis in Tetrahymena. Proceedings of the National Academy of Sciences *95*, 7480–7484.

Weiner, A., Lara-Astiaso, D., Krupalnik, V., Gafni, O., David, E., Winter, D. R., Hanna, J. H. and Amit, I. (2016). Co-ChIP enables genome-wide mapping of histone mark co-occurrence at single-molecule resolution. Nature Biotechnology *34*, 953–961.

Wu, H., Zeng, H., Dong, A., Li, F., He, H., Senisterra, G., Seitova, A., Duan, S., Brown, P. J., Vedadi, M., Arrowsmith, C. H. and Schapira, M. (2013). Structure of the Catalytic Domain of EZH2 Reveals Conformational Plasticity in Cofactor and Substrate Binding Sites and Explains Oncogenic Mutations. PLoS ONE *8*, e83737.

Xiang, M. and Li, S. (2013). Foxn4: A multi-faceted transcriptional regulator of cell fates in vertebrate development. Science China Life Sciences *56*, 985–993.

Yu, J.-R., Lee, C.-H., Oksuz, O., Stafford, J. M. and Reinberg, D. (2019). PRC2 is high maintenance. Genes & Development *33*, 903–935.

Yu, M., Mazor, T., Huang, H., Huang, H. T., Kathrein, K. L., Woo, A. J., Chouinard, C. R., Labadorf, A., Akie, T. E., Moran, T. B., Xie, H., Zacharek, S., Taniuchi, I., Roeder, R. G., Kim, C. F., Zon, L. I., Fraenkel, E. and Cantor, A. B. (2012). Direct Recruitment of Polycomb Repressive Complex 1 to Chromatin by Core Binding Transcription Factors. Molecular Cell *45*, 330–343.

Yuan, W., Wu, T., Fu, H., Dai, C., Wu, H., Liu, N., Li, X., Xu, M., Zhang, Z., Niu, T., Han, Z., Chai, J., Zhou, X. J., Gao, S. and Zhu, B. (2012). Dense Chromatin Activates Polycomb Repressive Complex 2 to Regulate H3 Lysine 27 Methylation. Science *337*, 971–975.

Zegerman, P., Canas, B., Pappin, D. and Kouzarides, T. (2002). Histone H3 Lysine 4 Methylation Disrupts Binding of Nucleosome Remodeling and Deacetylase (NuRD) Repressor Complex. Journal of Biological Chemistry *277*, 11621–11624.

Zentner, G. E., Tesar, P. J. and Scacheri, P. C. (2011). Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. Genome Research *21*, 1273–1283.

Zhang, P., Andrianakos, R., Yang, Y., Liu, C. and Lu, W. (2010). Kruppel-like Factor 4 (Klf4) Prevents Embryonic Stem (ES) Cell Differentiation by Regulating Nanog Gene Expression. Journal of Biological Chemistry *285*, 9180–9189.

Zhao, J., Sun, B. K., Erwin, J. A., Song, J.-J. and Lee, J. T. (2008). Polycomb Proteins Targeted by a Short Repeat RNA to the Mouse X Chromosome. Science *322*, 750–756.

Zhao, X. D., Han, X., Chew, J. L., Liu, J., Chiu, K. P., Choo, A., Orlov, Y. L., Sung, W. K., Shahab, A., Kuznetsov, V. A., Bourque, G., Oh, S., Ruan, Y., Ng, H. H. and Wei, C. L. (2007). Whole-Genome Mapping of Histone H3 Lys4 and 27 Trimethylations Reveals Distinct Genomic Compartments in Human Embryonic Stem Cells. Cell Stem Cell *1*, 286–298.

Zhou, W., Zhu, P., Wang, J., Pascual, G., Ohgi, K. A., Lozach, J., Glass, C. K. and Rosenfeld, M. G. (2008). Histone H2A Monoubiquitination Represses Transcription by Inhibiting RNA Polymerase II Transcriptional Elongation. Molecular Cell *29*, 69–80.

Zink, B. and Paro, R. (1989). In vivo binding pattern of a trans-regulator of homoeotic genes in Drosophila melanogaster. Nature *337*, 468–471.

Zylicz, J. J., Dietmann, S., Günesdogan, U., Hackett, J. A., Cougot, D., Lee, C. and Surani, M. A. (2015). Chromatin dynamics and the role of G9a in gene regulation and enhancer silencing during early mouse development. eLife *4*, 1–25.